https://ntrs.nasa.gov/search.jsp?R=19960020713 2020-06-16T05:10:26+00:00Z

NASA CR-199891

1995 Research Reports

NASA/ASEE Summer Faculty Fellowship Program

John F. Kennedy Space Center
and
University of Central Florida

P.14

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	November 1995	Contractor Report - Summer 1995

4. TITLE AND SUBTITLE

1995 Research Reports
NASA/ASEE Summer Faculty Fellowship Program

5. FUNDING NUMBERS

NASA Grant NGT-60002
Supplement: 19

6. AUTHOR(S)

See attached list

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Central Florida
Orlando, Florida 32816-2450
John F. Kennedy Space Center
Kennedy Space Center, Florida 32899

8. PERFORMING ORGANIZATION REPORT NUMBER

NASA CR-199891

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

National Aeronautics and Space Administration
Washington, D.C. 20546

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Unclassified - Unlimited
Subject Category 99

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This document is a collection of technical reports on research conducted by the participants in the 1995 NASA/ASEE Summer Faculty Fellowship Program at the Kennedy Space Center (KSC). This was the eleventh year that a NASA/ASEE program has been conducted at KSC. The 1995 program was administered by the University of Central Florida in cooperation with KSC. The program was operated under the auspices of the American Society for Engineering Education (ASEE) with sponsorship and funding from the Office of Educational Affairs, NASA Headquarters, Washington, D.C. The KSC Program was one of nine such Aeronautics and Space Research Programs funded by NASA Headquarters in 1995. The NASA/ASEE Program is intended to be a two-year program to allow in-depth research by the University faculty member. The editors of this document were responsible for selecting appropriately qualified faculty to address some of the many problems of current interest to NASA/KSC.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Research and Technology	640
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to ***stay within the lines*** to meet ***optical scanning requirements***.

Block 1. Agency Use Only *(Leave blank)*.

Block 2. Report Date. Full publication date including day, month, and year, if available (e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C	- Contract	**PR**	- Project
G	- Grant	**TA**	- Task
PE	- Program Element	**WU**	- Work Unit Accession No.

Block 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es). Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. Sponsoring/Monitoring Agency Name(s) and Address(es). Self-explanatory.

Block 10. Sponsoring/Monitoring Agency Report Number. *(If known)*

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with...; Trans. of...; To be published in.... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availability Statement. Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

- **DOD** - See DoDD 5230.24, "Distribution Statements on Technical Documents."
- **DOE** - See authorities.
- **NASA** - See Handbook NHB 2200.2.
- **NTIS** - Leave blank.

Block 12b. Distribution Code.

- **DOD** - Leave blank.
- **DOE** - Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.
- **NASA** - Leave blank.
- **NTIS** - Leave blank.

Block 13. Abstract. include a brief *(Maximum 200 words)* factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code *(NTIS only)*.

Blocks 17. - 19. Security Classifications. Self-explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.

1995 RESEARCH REPORTS

NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

JOHN F. KENNEDY SPACE CENTER

UNIVERSITY OF CENTRAL FLORIDA

EDITORS:

Dr. E. Ramon Hosler, University Program Director
Department of Mechanical and Aerospace Engineering
College of Engineering
University of Central Florida

Mr. Gregg Buckingham, NASA/KSC Program Director
Administration Office
John F. Kennedy Space Center

PREPARED FOR:

John F. Kennedy Space Center
Merritt Island, Florida

NASA Grant NGT-60002 Supplement: 19

Contractor Report No. CR-199891

October 1995

PREFACE

This document is a collection of technical reports on research conducted by the participants in the 1995 NASA/ASEE Summer Faculty Fellowship Program at Kennedy Space Center (KSC). This was the eleventh year that a NASA/ASEE program has been conducted at KSC. The 1995 program was administered by the University of Central Florida (UCF) in cooperation with KSC. The program was operated under the auspices of the American Society for Engineering Education (ASEE) with sponsorship and funding from the Office of Educational Affairs, NASA Headquarters, Washington, D.C. The KSC program was one of nine such Aeronautics and Space Research Programs funded by NASA Headquarters in 1995. The basic common objectives of the NASA/ASEE Summer Faculty Fellowship Program are:

a. To further the professional knowledge of qualified engineering and science faculty members;

b. To stimulate an exchange of ideas between participants and NASA;

c. To enrich and refresh the research and teaching activities of participants' institutions; and,

d. To contribute to the research objectives of the NASA centers.

The KSC Faculty Fellows spent ten weeks (June 1 through August 9, 1995) working with NASA scientists and engineers on research of mutual interest to the University faculty member and the NASA colleague. The editors of this document were responsible for selecting appropriately qualified faculty to address some of the many problems of current interest to NASA/KSC. A separate document reports on the administrative aspects of the 1995 program. The NASA/ASEE program is intended to be a two-year program to allow in-depth research by the University faculty member. In many cases a faculty member has developed a close working relationship with a particular NASA group that has provided funding beyond the two-year limit.

TABLE OF CONTENTS

1995 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

5,-32
7741
1-28

DIGITAL MODULATION SCHEMES FOR HIGH SPEED TRANSMISSION THROUGH LOW BANDWIDTH LOWPASS ANALOG LINKS

Dr. Alfred S. Andrawis, Assistant Professor
Mr. Quentin J. Flippin, Student
Electrical Engineering Department
South Dakota State University
Brookings, South Dakota

KSC Colleague - Po T. Huang
Communications/Fiber Optics

Contract Number NASA-NGT-60002
Supplement 19

July 22, 1995

ACKNOWLEDGMENTS

The authors are thankful for the NASA/ASEE Summer Faculty Program and for the opportunity to participate in it. The program was well organized and ran smoothly due to the efforts of Ray Hosler, Program Director, and Ms. Kari Stiles, Administrative Assistant. Ray and Kari did a great job at assisting, educating, and entertaining the program's participants through the many presentations, tours, and social events.

The authors would like to thank everyone in the Communications Division for accommodating and tolerating the four program participants that invaded their department. **Perry Rodgers, Larry Hand, Robert Stute, Sandra Gregory,** and **John Kassack** were all very friendly and extremely helpful.

The authors would like to extend special thanks to **Po T. Huang,** our NASA colleague. He assisted us with his vast experience and knowledge of the existing communication systems at KSC. Po also provided a personal tour and documentation pertaining to the research project.

The idea for this project was the brainchild of **Houston Galloway,** from INET. Houston also provided much helpful technical advice and insight pertaining to the existing fiber links.

A sincere thanks also goes to **Bill Jones,** from Advanced Systems, for loaning the authors the computer on which this report was typed.

Last but not least, the authors extend their gratitude to their families. Thanks goes to **Madaline, David, Mary and Danny Andrawis** for everything that they have done this summer to help make the authors' stay in Florida more enjoyable. Thanks also goes to **Jim, Carolyn, and Amanda Flippin** for all they have done in supporting this project.

DIGITAL MODULATION SCHEMES FOR HIGH SPEED TRANSMISSION THROUGH LOW BANDWIDTH LOWPASS ANALOG LINKS

Alfred S. Andrawis
and
Quentin J. Flippin

ABSTRACT

Existing NTSC standard will be phased out and replaced with HDTV standard within the next 10 years. Accordingly, the existing video network system operated by NASA will become obsolete and requires either replacement or modification to accommodate digital transmission. Network replacement is extremely expensive, hence, several digital modulation schemes are investigated in this report to accomplish digital transmission over existing analog links saving NASA from the cost of network replacement. There are two competing transmission systems available for HDTV transmission over limited bandwidth channels. The cost and performance of the two competing schemes are remarkably similar. However, the input data rate in such a case is limited to 40 Mbit/s. Transmission of higher data rates is possible using simple signal processing techniques. On the other hand, a third transmission system, multilevel pulse amplitude modulation (M-PAM) is proposed. M-PAM is the first stage of the well known M-VSB. This M-PAM scheme is much simpler and uses the channel more efficiently. The three schemes are compared and preliminary conclusions were made. Despite of several similarities, each modulation scheme has it unique merits. To determine the suitability of each scheme, more investigations and laboratory tests for all schemes are needed.

TABLE OF CONTENTS

LIST OF FIGURES

DIGITAL MODULATION SCHEMES FOR HIGH SPEED TRANSMISSION THROUGH LOW BANDWIDTH LOWPASS ANALOG LINKS

1. INTRODUCTION

1.1. Objective

To study present digital transmission technologies and their suitability for the new High Definition Television (HDTV) digital signals. Determine if the 12 MHz baseband of present video network can be economically utilized to transmit high data rates and save KSC from the expense of development and deployment of new network.

1.2. Motivations

The impact of HDTV will greatly enhance NASA's scientific, engineering and operational effectiveness. This includes improved resolution for TV images enhancing research, and technology activities, computational fluid dynamics, flow visualization, and material analysis. Moreover, NASA's requirement to disseminate information and video generated video to the public necessitates the prompt response to TV technology advances.

Video networks operated by NASA in general and KSC in particular will become obsolete within the next ten years. Several digital modulation techniques for terrestrial and cable broadcast have been proposed by HDTV Grand Alliance and other organizations. Unfortunately, none of these digital techniques is directly usable with the existing video network system operated by NASA. However, with some minor modifications, off-the-shelf components developed for terrestrial and cable digital broadcast could be utilized for digital transmission over the existing KSC analog links.

1.3. Scope

Recent advances in video compression techniques have spurred interest in the idea of digital television. Hence, the television industry is taking major steps toward a total digital television environment for both the existing National Television System Committee (NTSC) standard [1] and the new High Definition Television (HDTV) standard [2].

The Federal Communications Commission (FCC) has approved and tested the new, total digital, HDTV standard [3-5]. The FCC stated that the existing NTSC standard will be phased out and replaced with HDTV standard within the next 10 years. Accordingly, the existing video network system operated by NASA will become obsolete and requires either replacement or modification to accommodate digital transmission.

The Moving Picture Experts Group (MPEG) developed the MPEG-2 Video Standard which specifies the coded bit streams for high-quality digital video. MPEG-2 defines

three profiles and three levels of picture complexity. The main profile has a low-level resolution of up to 288 lines/frame (VHS tape resolution), 30 frames/sec, 2.53 million pixels/sec and 4 Mbit/s. The main level has up to 576 lines/frame (NTSC studio quality resolution), 30 frames/sec, 10.4 million pixels/sec and 15 Mbit/s. The high level has up to 1125 lines/frame (high definition television resolution), 60 frames/sec, 62.7 million pixels/sec and 60 Mbit/s.

United States Advanced Television Systems Committee (ATSC; a subcommittee of the FCC,) has approved Digital Television Standard for High Definition Television Transmission on April 1995 [3-5]. Uncompressed Digital Television Standard HDTV bit rate is 1.2 Gbit/s. The Digital Television Standard describes a system designed to transmit high quality video, audio, and ancillary data over a single 6 MHz channel. This means that encoding a video source whose resolution can be five times that of conventional television (NTSC) resolution requires a bit rate reduction factor of 50 or higher. To achieve this bit reduction, the system is designed to be efficient in utilizing available channel capacity by exploiting video and audio compression scheme based on MPEG-2 syntax. The allowable values for the field bit-rate-value are application dependent. In the primary application of terrestrial broadcast, this field shall correspond to a bit rate which is less than or equal to 19.4 Mbit/s. In the case of cable systems the corresponding bit rate shall be less than or equal to 38.8 Mbit/s.

ATSC also approved 8 VSB (vestigial side band modulation with 8 discrete amplitude levels) for terrestrial broadcast in standard NTSC 6 MHz bandwidth. Furthermore, 16 VSB (vestigial side band modulation with 16 discrete amplitude levels) is approved for the high data rate cable mode. Other digital modulation techniques have been extensively investigated and tested for terrestrial and cable HDTV broadcast transmission [6-11]. Several vendors developed chips, chip sets, and evaluation kits to achieve digital transmission for terrestrial and cable broadcast applications.

2. PRESENT SITUATION

Kennedy Space Center (KSC) is currently utilizing thousands of fiber-optic links for video transmission. The fiber optic-links consist of the following three basic units:1) transmitter (5000TX-KSC); 2) fiber; 3) receiver (5000 RX-KSC). The transmitters and receivers are manufactured by Video Products Group (formerly known as PCO).

2.1 Communications Network Layout

These links are combined to produce a communication system that is primarily used to transmit, high quality, analog television from the launch pad areas to the appropriate control rooms and other areas. There are around eighty cameras for each of the center's two launch pads. The video signals are transmitted via TV-39 cable from the camera to a room under the launch pad, the Pad Terminal Connection Room (PTCR). From here the electronic signals are converted to optical signals and sent to the Launch Control Center (LCC). It is here that the quality of the signals can be monitored and the video signals are recorded. From the LCC the signals are sent to the Communication Distribution Switching Center (CDSC). Then the CDSC the

signals are dispersed to the rest of the KSC. The distance that each link expands is relatively small.

2.2 Transmitter and Receiver Specifics

The signal that is to be transmitted has a nominal bandwidth of approximately 8MHz. To prepare this signal for transmission, it is passed through a pulse frequency modulator (PFM). The PFM's usable baseband bandwidth (3dB bandwidth) had been specified to be 12 MHz or greater [12]. The IF bandwidth is 32 MHz or greater. The signal is then transmitted using either a Light Emitting Diode (LED) or a Laser Diode (LD) through fiber. Either multimode or single mode fiber is used. The signal is then received and demodulated. The overall system has a baseband carrier-to-noise ratio (CNR) of 50dB or greater. Figure 1 shows a block diagram of the transmitter and receiver in some detail.

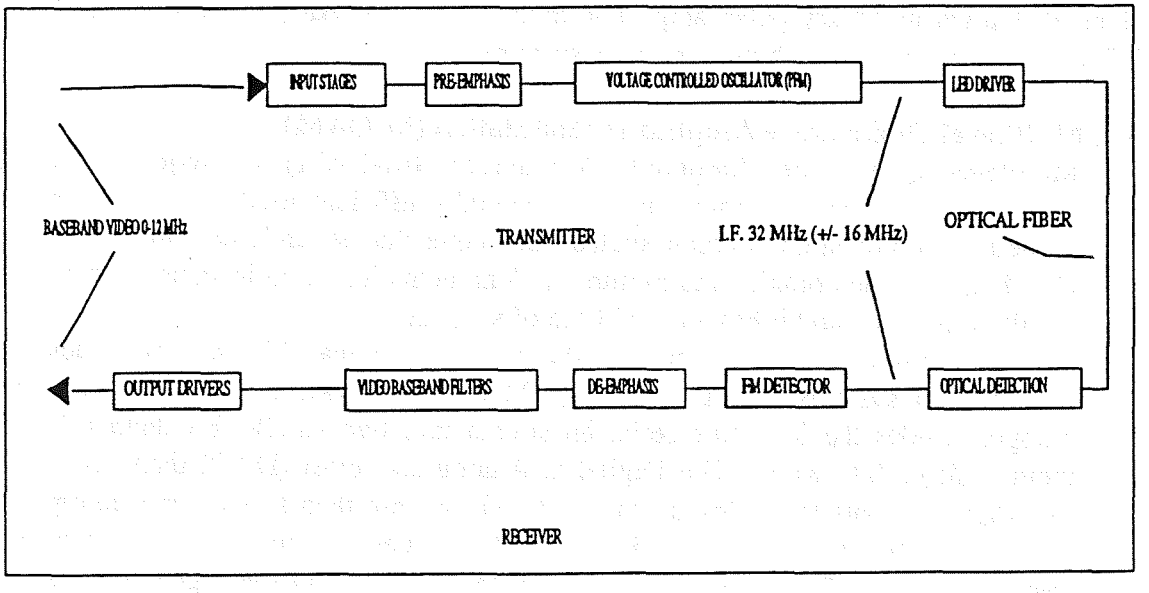


Figure 1: Block Diagram of Existing Systems

2.3 Investment

KSC has a considerable investment in the thousands of existing fiber-optic links. Each transmitter/receiver pair is purchased for approximately 5000 dollars. The lengths of fiber for each link range from one to seven km. The cost of the fiber (labor included) is estimated at twenty cents per meter. The initial cost of a link is estimated to range from $5200 to $6400.

2.4 Reliability

The existing system has been proven to be reliable. This particular equipment has been operating for 10 years with a minimum amount of failures. The majority of failures have been LED or ILD failures. Many of the optical fiber lines have been operating for 15 years without incident. The existing old copper links are still being updated to these fiber-optic links. Eventually the copper wire links will be phased out.

3. PROPOSED SOLUTIONS

Digital transmission over fiber is accomplished usually using base-band, on-off keying modulation, i.e. by switching the light-source on and off. However, long-haul digital transmission, where the spectrum is scarce, requires continuous wave (CW) modulation to generate a band-pass signal suited to the transmission medium- be it radio, cable, or hybrid fiber-optic/coaxial networks. In CW modulation [13], a digital signal can modulate the amplitude, or phase of a sinusoidal carrier wave. There are two competing band-pass modulation schemes for HDTV digital transmission systems: quadrature amplitude modulation (QAM), and vestigial side band modulation (VSB) [6]. Both QAM and VSB are bandwidth efficient, and have the same bandwidth efficiency. In this section, several conceptual and technical design considerations for the two competing transport schemes are presented. However, the KSC application is much simpler than what both of these two schemes are designed to do. A transmission scheme, uses a lowpass transmission channel, known as M-ary pulse amplitude modulation (M-PAM) is presented in Section 3.3 as an alternative for both two previous systems

3.1. Multilevel Quadrature Amplitude Modulation (M-QAM)

Multilevel Quadrature Amplitude Modulation (M-QAM) techniques have been generally developed for implementing spectrally efficient modems used in a band-limited communication systems, such as telephone channel and microwave radio links. QAM is, a public domain, and mature modem technology requiring no license fees or royalty payments and it has a broad base of support.

The basic principal of quadrature modulation is to increase the spectral efficiency in transmission systems. In the basic QAM system shown in Figure 2, the symbol mapper divides the incoming serial bit stream into two parallel bit steams. Each od them is $M/2$ bits wide. The Digital-to-Analog converter (DAC) then converts the two digital signals to analog signals. The signals are then filtered and multiplied by two perpendicular sinusoidal signals with frequency equal to the intermediate frequency (IF). The sum of the two signals is a M-QAM signal with spectral efficiency r_b/B_T equal to:

$$r_b/B_T \approx \log_2 M \qquad \text{bit/s/Hz} \tag{1}$$

where r_b is the data bit rate, B_T is the transmission bandwidth, and M is the number of M-ary levels. Hence, for 256 QAM, $r_b/B_T = 8$. At the demodulator side the reverse operation is performed to retrieve the original transmitted data stream. Figure 3 illustrates the IF I-Q Constellation for QAM signal with M=16. Many attractive features can be found when this scheme is applied to HDTV signal transport.

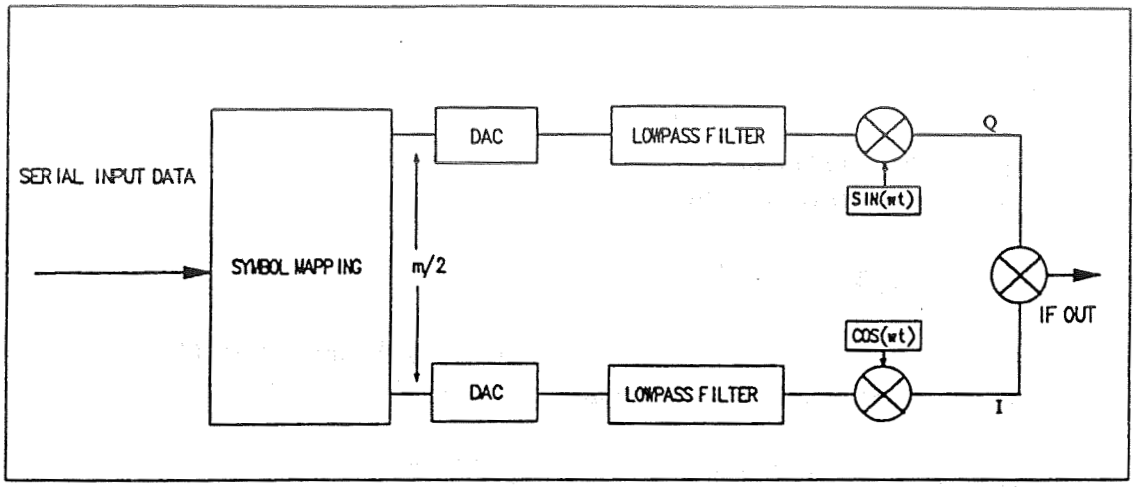


Figure 2: Block Diagram of Basic QAM

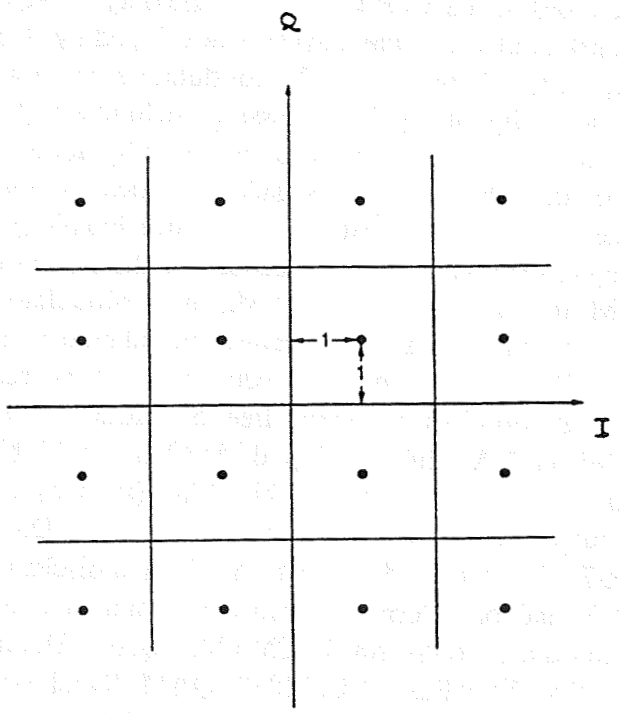


Figure 3: 16-QAM IF I-Q Constellation

The first feature is that this transport system can be economically implemented, has been tested, and proved its reliability. Recently, proponents of this modulation scheme produced reasonably priced VLSI modem chips [14-18].

The second feature is that, by using multilevel modulation, the efficient utilization of bandwidth becomes possible and, consequently, existing KSC analog links can carry relatively high bit rates. For 256 QAM, bandwidth efficiency is 8 bits/Hz, i.e. the

maximum bit rate to be transmitted via 11 MHz of channel bandwidth theoretically is 88 Mbit/s.
The third feature is that multilevel QAM makes it possible to accept parallel bits more easily. The number of parallel bits to be accepted by the modulator as a function of the M-ary levels is given by:

$$m = \log_2 M \quad \text{bit/s/Hz} \tag{2}$$

where m is the number of parallel bits and M is the number of M-ary levels. Hence, for 256 QAM modulation the number of parallel bits to be handled is 8 bits.

3.1.1. Vendor Implementation of M-QAM Technology

Major proponents of QAM technology are Broadcom Corp., AT&T, Scientific-Atlanta, Samsung, Hewlett-Packard, and General Instruments [19]. To the best of authors' knowledge, the only 256-QAM modulator-demodulator chip set available in the market is developed by Broadcom (QAMLink™). The complexity of typical QAM modulators is greatly reduced by combining advanced digital signal-processing techniques (DSP) and high-speed VLSI fabrication. Broadcom's design elegantly removes the typical DACs from inside the chip to the outside significantly reducing its complexity and consequently its cost. Hence, using Broadcom's QAM modulator requires only a single, external DAC chip instead of the typical two DAC's used in classical QAM modulators. The QAMLink™ Modulator chip BCM3023 provides baseband filtering and quadrature modulation of a QAM signal at which the digital IF center frequency is equal to the baud rate [19]. Theoretically, 8 bit DAC is sufficient for error free operation ($2^8 = 256$ levels), however, the QAMLink™ Modulator chip BCM3023 has 12 bit outputs. For normal bit-error rate requirement, 10 bit DAC (such as Analog Devices AD9721) is adequate. For lower bit-error rates, 12 bit DAC (such as Analog Devices AD9713) is required. Due to a joint nondisclosure agreement between KSC and Broadcom Corp., the block diagram of actual QAMLink™ Modulator-Demodulator chip set (BCM3020 QAM Demodulator, BCM3021 QAM Adaptive Equalizer, BCM3022 QAM Synchronization IC, and BCM3023 QAM Modulator) can not be provided here for publication. However, all QAMLink™ Proprietary Information is available to interested individuals working with the KSC Communications Group.

Broadcom also manufactures more integrated chips such as the QAMLink receiver (BCM3100) and the QAMLink Dual-Channel Receiver with Forward Error Correction (BCM3115).
The BCM3100 integrates the BCM3020 QAM Demodulator, BCM3021 QAM Adaptive Equalizer, and BCM3022 QAM functions into one chip. It is a complete single chip digital IF receiver supporting programmable QPSK, 16, 64 and 256 QAM transmission. Moreover, the BCM3115 is a complete single chip digital IF receiver which accepts a high bandwidth (40 Mbit/s maximum)

in-band data stream and low bandwidth (2 Mbit/s maximum) out-of-band data stream and delivers two demodulated, error corrected output data streams.

3.2. Multilevel Digital Vestigial Side band Modulation (M-VSB)

This Digital Vestigial Side Band Modulation (M-VSB) method (shown in Figure 4) applies Nyquist pulse shaping to a polar input signal producing a band limited modulating signal with a bandwidth B given by [13]

$$B = (r/2) + \beta_N \tag{3}$$

where r is the symbol rate, and β_N is the rolloff factor. For theoretical brick-wall filters $\beta_N = 0$ (0 % rolloff), while $\beta_N = r/2$ for the spectrum with smooth rolloff filters (100% rolloff). The Symbol rate r can be expressed in turns of the data bit rate as

$$r = r_b/\log_2 M \tag{4}$$

where r_b is the data bit rate, and M is number of M-ary levels. The VSB filter then removes all but a **vestige** of width β_V from one side band so the transmitted signal looks like Figure 5, a band limited spectrum with bandwidth B_T where

$$B_T = (r/2) + \beta_N + \beta_V \tag{5}$$

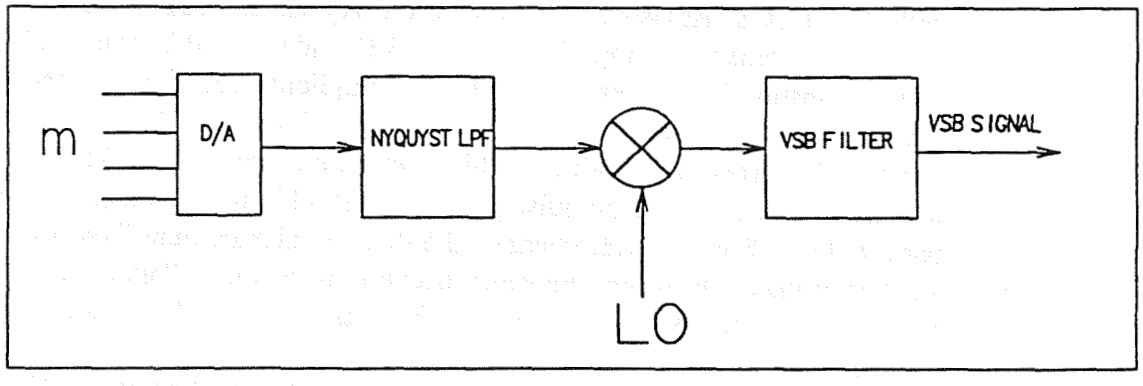


Figure 4: Block Diagram of Basic VSB Modulation

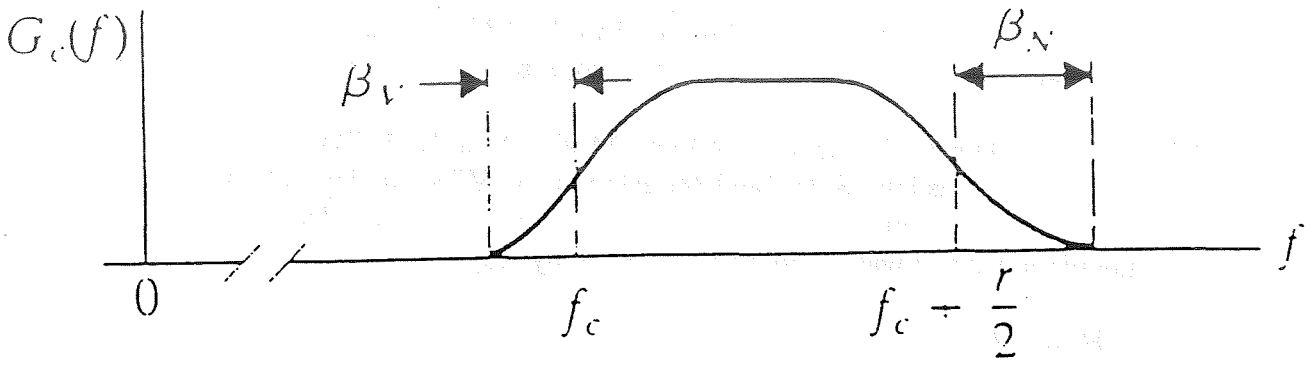


Figure 5: Digital VSB Power Spectrum

Hence, the upper limit for the spectral efficiency r_b/B_T is equal to:

$$r_b/B_T = 2\log_2 M \quad \text{bit/s/Hz} \tag{6}$$

This upper bound holds when $\beta_N \ll r$ and $\beta_V \ll r$.

3.2.1. *Vendor Implementation of M-VSB Technology*

Major proponents of VSB technology are Zenith, Philips, Compression Labs Inc.(CLI), and HDTV Grand Alliance. The Grand Alliance, in cooperation with the FCC's Advisory Committee on Advanced Television (ACATS), concluded extensive testing of Zenith's "VSB" and General Instrument "QAM" digital transmission systems. Based on simplicity, ruggedness, and Grand Alliance recommendation, ACATS approved Zenith's VSB system as the standard for terrestrial broadcast of high-definition television (HDTV).

M-VSB method applies Nyquist pulse shaping [13] to a symbol train with a bandwidth r. For a 6 MHz channel, the theoretical maximum (Nyquist) rate is 12 Msymbol/s but is unachievable because it requires "brick-wall" filters. Using surface acoustic wave (SAW) filter technology, 10.76 Msymbol/s is economically feasible [20]. At this symbol rate, if binary symbols are transmitted the maximum data rate to be transmitted will be 10.76 Mbit/s. If 4-level symbols are sent, the maximum data rate to be transmitted is 21.5 Mbit/s. For 8-level and 16-level symbols, the maximum data rate to be transmitted is 32.3 Mbit/s and 43 Mbit/s respectively. Zenith VSB modem is actually designed to handle 2-, 4-, 8-, and 16-level VSB modes. The VSB receiver will be able to automatically lock with modes under head end control. The IF frequency response of Zenith's 16-VSB can be seen in Figure 6.

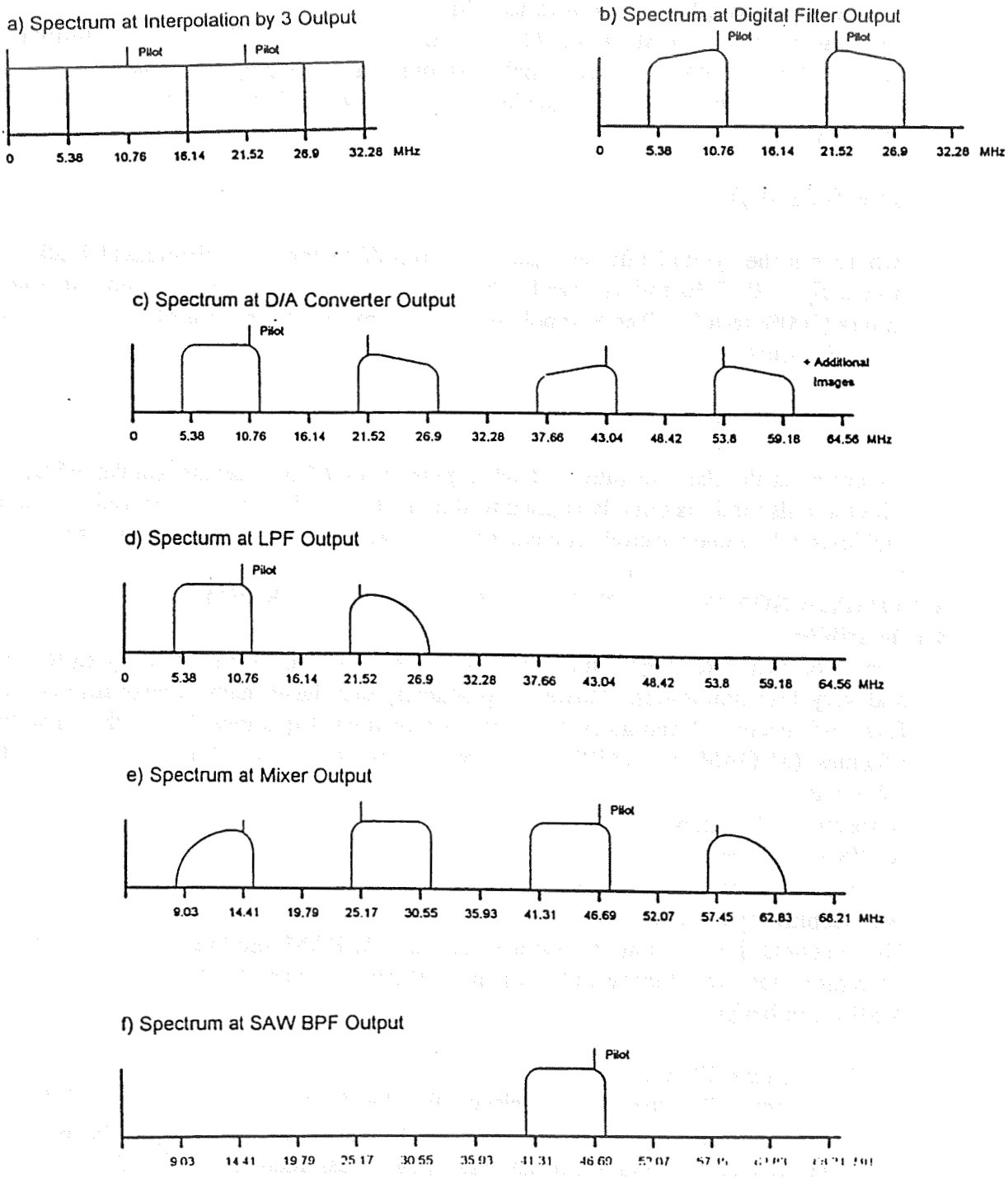


Figure 6: Zenith's 16-VSB Modulator Frequency Spectrum

3.3. M-ary Pulse Amplitude Modulation (M-PAM)

Similar to the first stage of M-VSB modulation, this digital Pulse Amplitude Modulation (M-PAM) method applies Nyquist pulse shaping to a polar input signal producing a band limited modulating signal with a bandwidth B given by equation (3) [13]

$$B = (r/2) + \beta_N$$

where r is the symbol rate, and β_N is the rolloff factor. For theoretical brick-wall filters $\beta_N = 0$ (0 % rolloff), while $\beta_N = r/2$ for the spectrum with smooth rolloff filters (100% rolloff). The Symbol rate r can be expressed in turns of the data bit rate given in equation 4 as

$$r = r_b/\log_2 M$$

where r_b is the data bit rate, and M is number of M-ary levels. Unlike VSB, this filtered pulse train is directly applied to the input of the low pass transmission channel achieving the same spectral efficiency and performance as the two previous schemes.

4. COMPARISON OF PROPOSED TRANSMISSION SYSTEMS

4.1. Feasibility Study

The three proposed digital transmission systems theoretically have many differences and very few similarities. However, practically they have many similarities and very few differences. Laboratory tests and simulations [6] showed that the first two schemes (M-QAM and MVSB) have very close behavior (almost similar) for the following:

- spectrum efficiency,
- C/N vs. bit error rate,
- carrier recovery (without inserting pilots), and
- susceptibility to jitter.

No practical information is available for the M-PAM modulation scheme in the literature, however, theory indicates that its' performance should be identical to M-VSB if not better.

4.1.1. *Spectrum Efficiency*

Theoretically, spectrum efficiency for 256-QAM, 16-VSB, and 16-PAM, as given in equations (1) and (6), are the same and it is equal to 8 bit/Hz. However, practical spectrum efficiency is less than ideal. In QAM systems, spectrum spill over limits the spectrum efficiency to ~7.2 bit/Hz. Similarly for VSB scheme, roll-off of both Nyquist filter and VSB filter limit the spectrum efficiency to the same as the QAM. Theory indicates that the spectrum efficiency for M-PAM is in the same range as the two other schemes if not little bit better.

4.1.2. *Minimum Required Carrier-to-Noise Ratio*

Bit error rate (BER) for any M-ary system is a function of the carrier-to-noise ratio (CNR) and the number of M-ary levels (M). According to theoretical and laboratory research, the behavior of all systems is identical [6]. Therefore, the performance of the QAM system is explored then, by induction, the performance of the M-VSB and M-PAM systems are estimated.
The symbol error rate of a QAM signal is [9] defined as

$$P_s(e) = 2(1 - \tfrac{1}{\sqrt{M}})Q\left(\sqrt{\tfrac{3}{M-1}(CNR)}\right) \tag{7}$$

where $Q(x)$ is defined as

$$Q(x) = \tfrac{1}{\sqrt{2\pi}}\int_x^{\infty} e^{-t^2/2}dt \tag{8}$$

In addition, if the multilevel signal is assumed to be Gray coded, as is conventionally adopted, the bit error rate $P_B(e)$ is given by

$$P_B(e) \approx \tfrac{P_s(e)}{\log_2\sqrt{M}} \tag{9}$$

From equation 7 and equation 9, the bit error rate $P_B(e)$ can be written as

$$P_B(e) = \tfrac{2}{\log_2\sqrt{M}}(1 - \tfrac{1}{\sqrt{M}})Q\left(\sqrt{\tfrac{3}{M-1}(CNR)}\right) \tag{10}$$

For bit error rate $P_B(e)$=10^{-9}, $M = 256$ levels and using equation (10), minimum CNR is then calculated to be equal to 34.6 dB.
Present analog TV has a CNR equal to, or better than 50 dB [12], Therefore, either system (256-QAM or 16 VSB) will achieve a bit error rate $P_B(e) \leq 10^{-9}$. Figure 7 shows the relationship between the probability of error vs. CNR for both QAM and VSB systems with and without forward error correction (FEC).

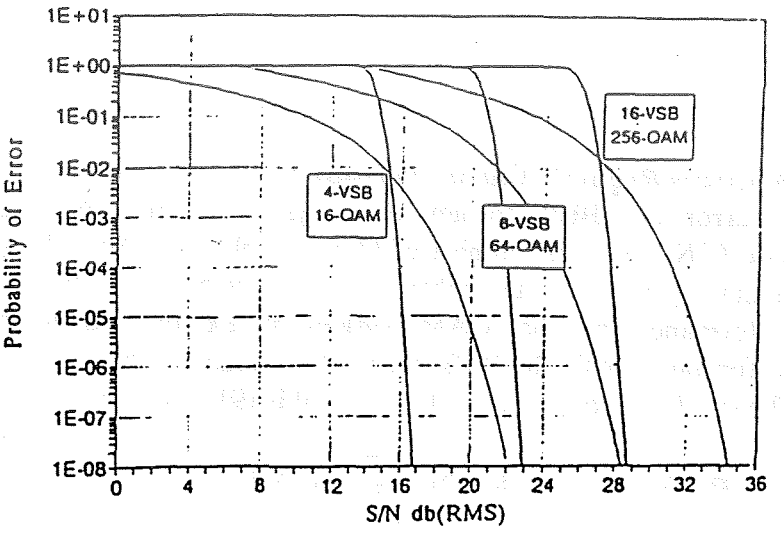


Figure 7: The probability of error vs. CNR for both QAM and VSB systems with and without forward error correction (FEC)

4.1.3. *Local Oscillator Frequency Selection*

With data bit rate of 40 Mbit/s, the upper corner frequency of the signal is less than 12 MHz for both Broadcom BCM3023 QAM Modulator (256 levels) or Zenith VSB modulator kit (before the mixer) [20]. **Therefore, no further frequency processing is needed.** However, the output of both evaluation kits, has an IF center frequency much higher than the upper 3-dB corner frequency of the channel. In such case, a frequency mixer is needed to shift the signal's spectrum down to PCO lowpass channel range (i.e. 10Hz-12MHz range.) Following is a procedure for the selection of a local oscillator frequency to shift the spectrum of Broadcom's Evaluation System down to lay within the lowpass bandwidth of the PCO transmission channel. Similar procedure is valid to shift the spectrum of Zenith's VSB Evaluation Kit.

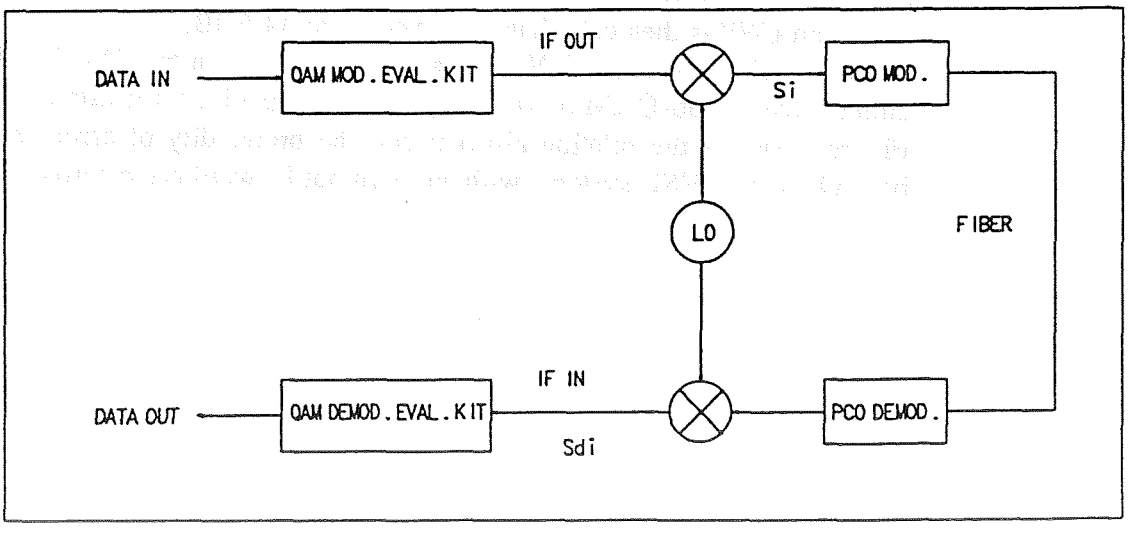


Figure 8: Block diagram of experiment setup using Broadcom BCM93100 QAMLink Development System

If the evaluation kit has an intermediate frequency (IF) of ω_{if} rad/s. The PCO input signal S_i is equal to

$$S_i = a(t)cos(\omega_{if}t + \gamma(t))cos\,(\omega_{lo}t)$$

$$= \frac{a(t)}{2}\{cos[(\omega_{if} - \omega_{lo})t + \gamma(t)] + cos[(\omega_{if} + \omega_{lo})t + \gamma(t)]\} \quad (11)$$

If $(\omega_{if} + \omega_{lo})$ is selected much larger than the upper 3-dB frequency of the transmission channel, then the signal $cos[(\omega_{if} + \omega_{lo})t + \gamma(t)]$ is filtered out. At the receiver side, the QAM demodulator input signal S_{di} is given by:

$$S_{di} = Ka(t)[cos(\omega_{if} - \omega_{lo})t + \gamma(t))cos\,\omega_{lo}t]$$

$$= \frac{Ka(t)}{2}[cos(\omega_{if}t + \gamma(t)) + cos((\omega_{if} - 2\omega_{lo})t + \gamma(t))] \quad (12)$$

where K is a constant (function of received signal level). To avoid aliasing the following condition should be observed:

$$\omega_{if} - (2\omega_{lo} - \omega_{if}) \geq 2\pi(BW)$$

$$\omega_{if} - \omega_{lo} \geq 2\pi(BW)/2 \quad (13)$$

Where BW is the bandwidth of the IF signal. For Broadcom BCM93100 QAMLink Development System, the IF frequency is 43.75 MHz, $BW = 11$MHz (for $\tau_b = 80$ Mbit/s with 256-QAM). Hence, if we select $f_{lo} = f_{if} - (BW/2)$, *i.e.* $f_{lo} = 38.25$ MHz, we find that the difference between these two frequencies $(f_{lo} - f_{if})$ is equal to 5.5 MHz, which is the center of the transmission band. While the distance between two adjacent images after the receiver mixer is $2f_{lo} = 76.5$ MHz, that is to say, no aliasing will occur at the receiver's side.

Differences between the two competing techniques and the associated advantages and disadvantages are listed in the following sections

4.2. Differences

4.2.1. *Carrier Presence*

One of the most significant differences is the fact the VSB do not need carrier while QAM needs two perpendicular carriers (M-VSB with no carrier converges to M-PAM.) This makes VSB more suitable for transmitting high data rates on lowpass channels. **However, for data rates limited to 40 Mbit/s, the IF signal, before the up-converter, fits directly within the lowpass transmission band.** Thus this point becomes irrelevant for low data rate transmission systems.

4.2.2 Compatibility with Future HDTV Set

The FCC's Advisory Committee on Advanced Television Service (ACATS) endorsed VSB technology for cable and terrestrial digital transmission. Therefore, it is more likely that consumer's HDTV sets will be equipped with VSB receiver and demodulator. On the other hand, in case of QAM transmission, both the receiver and the demodulator has to be provided to users as a set-top box.

4.2.3. ***Required CNR For Acquisition***

VSB as proposed by Zenith, transmits a pilot tone at the carrier frequency, which is in the roll-off of the transmitted VSB signal. The pilot is easily made by adding a DC bias at baseband. The pilot increases the transmit signal power by 0.3 dB, resulting in an effective signal-to-noise (SNR) loss of 0.3 dB. However, this pilot signal enables the acquisition of 16-VSB signal to be possible even at CNR = 0 dB compared to minimum CNR = 18 dB for 256-QAM.

4.2.4. ***Availability Of The Technology***

VLSI modem chips for both modulation techniques are readily available of-the-shelf manufactured by Broadcom and Zenith. However, if carrierless VSB is used (i.e. Digital M-PAM), all components needed are readily available from vendors other than the aforementioned two vendors. This may reduce the cost of the system.

5. RECOMMENDATIONS

The performance and inherent cost of QAM and VSB are remarkably similar for high speed digital transmission over analog links. Therefore, both modulation techniques should be investigated and their suitability for KSC networks should be carefully examined. This section proposed a plan consisting of three phases. These phases may be implemented in orderly fashion or more than one phase simultaneously.

Phase one concentrates on the transmission of relatively low data rates (40 Mbit/s or less). This rate is sufficient to transmit two 19.3 Mbit/s HDTV compressed signals or one distribution quality HDTV signal. Moreover, the first phase will give KSC engineers and scientists more understanding for vendor's VLSI chip design, pinout, and system design. During this first phase, more understanding for the suitability of each scheme will also be acquired. Transmission of higher data rates will be experimented during phase two. Data rates up to 80 Mbit/s 256-QAM with shifted carrier will be experimented on existing fiber-optic transmission systems. Baseband carrierless VSB (M-PAM) also should be simulated and practically tried with data rates higher than 40 Mbit/s. Selection of the most suitable scheme, design, experimentation and documentation of the selected modulator/ demodulator pair will be accomplished during this last proposed phase (phase three) of the project.

5.1. Phase I

- Procurement of the following evaluation systems:

System	Cost
- Broadcom BCM93100 QAMLink™ Development System;	$30,000
- Zenith VSB Modulation System Test Transmitter and Receiver	$9,800
- Miscellaneous (filters, mixers, connectors DACs, ADCs, etc.)	$5,200
TOTAL	**$45,000**

- Experiment in laboratory the transmission of 256-QAM IF signal with input data rate of 40 Mbit/s. The QAM center frequency and the bandwidth of the IF signal is within the limits of KSC low pass transmission band. The power frequency spectrum of KSC superimposed on the power spectrum of the QAM signal is shown in Figure 9.
- Experiment in laboratory the transmission of 16-VSB IF signal after the LPF and just before mixer as depicted in Figure 7 (d). The power frequency spectrum of KSC superimposed on the power spectrum of this VSB signal is shown in Figure 10.

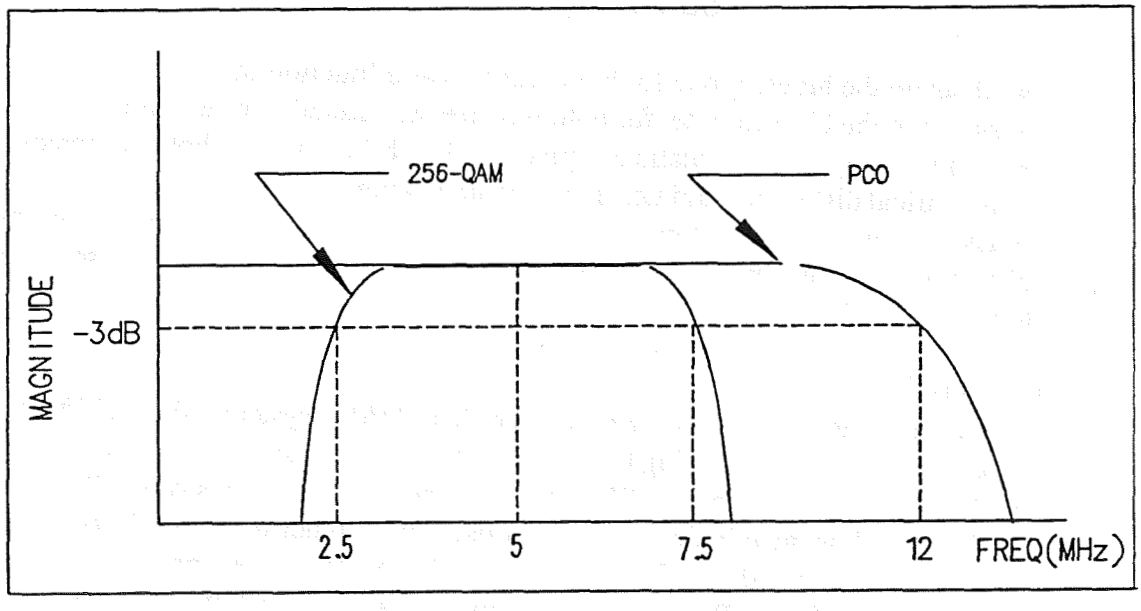


Figure 9: Power Frequency Spectrums of KSC Channel Freq. Response and 256-QAM With Input Data Rate of 40 Mbit/s

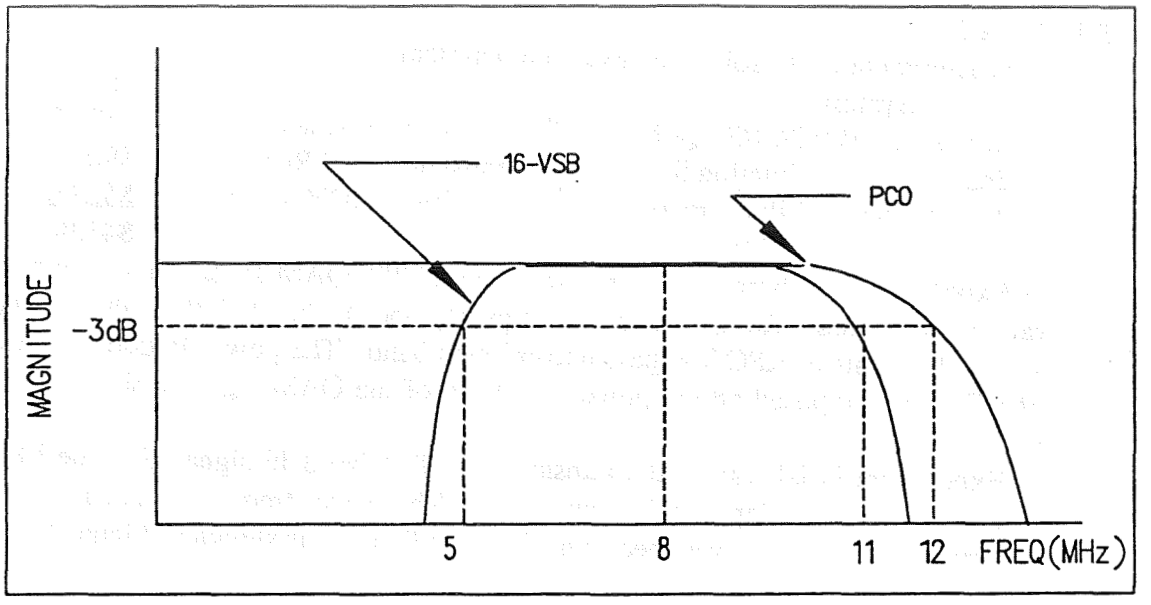


Figure 10: Power Frequency Spectrums of KSC Channel Freq. Response and 16-VSB With Input Data Rate of 40 Mbit/s

- Measure the bit error rate for both schemes as a function of CNR.
- Measure the bit error rate for both schemes as a function of the jitter.
- Using computer simulations, predict the behavior of both schemes under unsymetrical distortions and unsymetrical attenuation.
- Experiment in laboratory the behavior of both schemes under unsymetrical distortions and unsymetrical attenuation. Simulation results should be used as a guide to avoid any discrepancies.

5.2 Phase II

- Increase the input data rate for the 256 QAM system while shifting the IF spectrum to fit within existing low pass transmission band.
- Simulate and implement VSB base band transmission (carrierless VSB or Digital M-PAM as illustrated in Section 3.3) for data rates higher than 40 Mbit/s. Simulation will determine the order of filters required and expected system performance.
- BER vs. CNR and BER vs. jitter for high data rates for both schemes will also be experimented.
- Evaluate all schemes under consideration and disciminate results to other NASA centers.

5.3 Phase III

- Select the more suitable scheme or schemes. Most probably more than one scheme will be used depending on the specific application.
- Design and build a prototype for the selected modulation schemes.
- Test and build a prototype modulator/demodulator system using the selected scheme.
- Write documentation and maintenance manual for the system.

6. CONCLUSION

Either one of the two competing modulation schemes, M-QAM or M-VSB, could be used for transmitting digital signals over KSC's analog links. The cost and performance of the two competing schemes are remarkably similar. In the simplest form, the IF signal output from the modulator chip, could be transmitted with no further frequency shifting or processing. However, the input data rate in such a case is limited to 40 Mbit/s. Theoretically, KSC video communications network is capable of carrying bit rates up to 90 Mbit/s Therefore, vendor's VLSI chip implementation of the two schemes do not take full advantage of the video network of KSC, i.e. the transmission channel is not used the optimum way.

To use the transmission channel more efficiently shifting the IF signals to occupy the low pass frequency range, of both systems is essential. Better off, a simplified version of M-VSB (M-PAM) is proposed. This M-PAM scheme is much simpler and uses the channel more efficiently. With either shifted M-QAM IF carrier, or M-PAM, transmission of data rates close to the maximum possible rate is feasible.

Hands-on experience is needed for all modulation schemes to definitely rule out the suitability of each scheme. Therefore, the procurement of Broadcom's evaluation kit and Zenith's test board is recommended. Since there is no system available yet to evaluate the M-PAM scheme, a small extra funding is recommended to build a system for testing and evaluate the M-PAM scheme. This evaluation system could easily be built and tested here at KSC laboratories.

REFERENCES

[1] The Society of Motion Picture and Television Engineers, "Component Video Signal 4:2:2 Bit-Parallel Interface," SMPTE Standard, ANSI/SMPTE 125M-1992, JULY 1992.

[2] The Society of Motion Picture and Television Engineers, "Signal Parameters-1125/60 High-Definition Production System," SMPTE Standard, ANSI/SMPTE 240M-1988, March 14, 1988

[3] "Digital Television Standard for HDTV Transmission," ATSC Document,

[4] "Guide to the Use of the Digital Television Standard for HDTV Transmission," ATSC Document,

[5] "Digital Audio Compression (ac-3) Standard," ATSC Document,

[6] K.J. Kerpez, "A Comparison of QAM and VSB for Hybrid Fiber/Coax Digital Transmission," IEEE Transactions on Broadcasting, vol 41, March 1995

[7] M. Sablatash, "Transmission of All-Digital Advanced Television: State of the Art and Future Directions," IEEE Transactions on Broadcasting, vol 40, June 1994

[8] X. Lu, E. Bodeep, and T.E. Darcie, "Broad-Band AM-VSB/64 QAM Cable TV System Over Hybrid Fiber/Coax Network," IEEE Photonics Technology Letters, vol 7, April 1995

[9] N. Kanno and K. Ito, "Fiber-Optic Subcarrier Multiplexing Video Transport Employing Multilevel QAM," IEEE Journal on Selected Areas in Communications, vol 8, September 1990

[10] P. Fockens, "The 43 Mbit/s Digital 16-VSB Modem for Cable TV," Communications Technology, pp 41-54, May 1994

[11] L. Goldberg, "IC Opens 500-Channel Frontier to Cable System," Electronic Design, November 7, 1995

[12] Specifications of Wideband Fiber Optic Terminal Equipment, KSC document 79K33043, page 9, February 1991

[13] A. B. Carlson, "An Introduction to Signal and Noise in Electrical Communication," Communication Systems, McGraw-Hill Book Company, 1986

[14] H. Samueli, "QAM Modem Technology for Digital Cable TV," CED, December 1994

[15] R. Quinnel, "Building the Digital-Video Pipeline," EDN, pp 39-42, December 22, 1994

[16] R. Krause, "GI, Broadcom Merge QAM Techniques," Electronic News, Vol 41, June 5 1995

[17 L. Goldberg, "IC Opens 500-Channel Frontier to Cable Systems," Electronic Design, November 7, 1994

[18] L. Wirbel, "Broadcom Shrinks Three-Chip QAM to One," Electronic Engineering Times, Issue 824, November 21, 1994

[19] "QAMLink Modulator," Broadcom, Document BCM3023-SP3, March 16, 1995

[20]"Zenith VSB Modulator User Mannual," Zenith Electronics Corporation, V1.3, June 19. 1995

1995 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

52-54
7742
P. 26

ADVANCED PAYLOAD TRANSFER MEASUREMENT SYSTEM (APT-MS) MECHANICAL FEATURES

Dr. William V. Brewer
Professor
Technology Department
Jackson State University
Jackson, Mississippi

KSC Colleague - Eduardo Lopez del Castillo
Robotics

Contract Number NASA-NGT-60002
Supplement 19

August 1, 1995

PAGE 24 INTENTIONALLY BLANK

ACKNOWLEDGEMENTS

Eduardo Lopez provided a good work environment upon arrival here at KSC. He framed the problem well and got the project off to a good start. Rich Bennett shared his design drawings with me in a format that my equipment can use. I enjoyed working with both Eduardo and Rich on the APT system development.

My appreciation goes to these and many others who made the work here a pleasure.

My thanks also to Ramon Hosler and Kari Styles for putting up with us, arranging all of the extra curricular activities, and answering all the same questions over and over.

It's been fun guys and I hope we can do it again sometime.

ABSTRACT

Purpose: APT is a relatively inexpensive electro-mechanical device that reduces both time and manpower required to make shuttle payload transfers.

Objective: Explore possible mechanical sources of measurement error.
Develop an improved prototype design that is compact, inexpensive, and addresses the measurement error issues.

Motivation: During the last two feet or so of transfer, for a large or heavy payload to its restraining fixture, the consequences of unplanned contact (or impact) between payload and support structure can range from merely annoying to something approaching disaster.. Current transfer methods employ technicians with meter sticks stationed at the various hold-down locations to estimate the distances to contact. This information is communicated orally to the crane operator. It is understandable that this work proceeds carefully and therefore slowly. APT measurement system would provide a GUI for the "move-conductor" (crane operator) so he could see the displacements of all hold-down interfaces as they move together and thus have a more accurate, comprehensive and "real-time" picture of the engagement activity.

Accomplishments:
An error model attempts to include all estimatable sources of mechanical error.
Design features were introduced to reduce or eliminate major sources of error.

SUMMARY

Purpose: APT is a relatively inexpensive electro-mechanical device that reduces both time and manpower required to make shuttle payload transfers.

Objective: Explore possible mechanical sources of measurement error.
Develop an improved prototype design that is compact, inexpensive, and addresses the measurement error issues.

Description: APT has a spring loaded retractable cable that is attached to a moving object whose location must be monitored. The drum (or reel) on which the cable is stored is gimbled and instrumented to supply the spherical coordinates of the moving object relative to the APT device. These may be converted to rectangular coordinates or whatever form best facilitates a useful GUI for the operator trying to guide the moving object.

Motivation: During the last two feet or so of transfer for a large or heavy payload to its restraining fixture, it is difficult to avoid undesirable contact (or impact) between payload and support structure. Current transfer methods employ technicians with meter sticks to estimate the distances to hold-down locations. This information is communicated orally to the crane operator. It is understandable that this work proceeds carefully and therefore slowly. APT measurement system would provide a GUI for the "move conductor" (crane operator) so he could see the displacements of all hold-down interfaces as they move together and thus have a more accurate, comprehensive and "real-time" picture of the engagement activity.

Accomplishments: An error model attempts to include all estimatable sources of mechanical error which fall into two categories: cable guide errors, bearing friction errors.

Estimates of positional error for the first prototype ranged from .5" to nearly 3" as the radial distance approached the limit of 72" for the device. Since the goal is an error on the order of .050" at a radial distance of 24", this is clearly unacceptable.

Cable guide errors could account for as much as 2/3 of the error, some effort was expended designing a better one. The so-called Zero Clearance Cable Guide was proposed as a possible solution to this part of the problem.

Bearing Friction is difficult to model. We have drawn from several sources to put together a model with as much of the geometric characteristics of the bearings and their operating environment as are available rather than relying on empirical aggregated numbers for specific bearings.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

I

INTRODUCTION

1.1 HISTORY

APT-MS is the brain child of Eduardo Lopez. Concept investigation of mechanical aspects was initiated 6/94 by N.S.Malladi [1], followed by work of E.Lopez and R.Bennett since then [2,3]. During that time an earlier hardware prototype was built and tested [2], primarily to develop the electronics associated with reading of angular displacement sensors. The electronic work was done by others in this multi-disciplinary effort. Electronic aspects are indispensable to the results of the project, but are beyond the scope of this discussion of mechanical functions.

1.2 MOTIVATION

Undesirable contact (or impact) between a large, cumbersome payload and its restraining fixture is difficult to avoid during the last 2ft. of transfer (Figure 1-1). Current transfer methods employ technicians with meter sticks to estimate the distances to hold-down locations. This information is communicated orally to the crane operator. It is understandable that this work proceeds carefully and therefore slowly. APT measurement system provides a Graphic User Interface (GUI), shown in Figure 1-1, lower left. A "move-conductor" can see displacements of all hold-down interfaces as they move together and thus have a more accurate, comprehensive and "real-time" picture of engagement activity. Optical devices are an alternative shown in Figure 1-1, lower middle, but these may be more expensive and require greater stand-off than the mechanical APT system discussed here.

1.3 DESCRIPTION

APT has a spring loaded retractable cable that is attached to a moving object whose location must be monitored (Figure 1-1, lower right). The drum (or reel) on which the cable is stored is gimbled and instrumented to supply spherical coordinates for the moving object relative to an APT device. These may be converted to rectangular coordinates or whatever form best facilitates a useful GUI for an operator trying to guide the moving object.

Figure 1-2 shows the cable drum or spool nested inside a cable guide cage or pointer. These elements interact but rotate on separate horizontal bearing shafts: cable shaft on the right; cage on the left. These in turn are carried in the uprights of a "Y" shaped yoke whose lower stem shaft pivots about a vertical axis on bearings set in the fixed base. At the outer end of each shaft is an optical encoder to read one of the spherical coordinates.. The base mounted encoder measures "yaw" angular displacement of the yoke stem code wheel. On the left, a yoke mounted encoder measures the "pitch" angle of the cage shaft and therefore of the cable-exit, guide-tube. On the right, a yoke mounted encoder measures rotation of the cable storage drum shaft. Radial coordinate measurements subtract cage rotation from drum rotation and convert the difference to length of cable played out. Cable retraction is accomplished with a "clock" spring mounted between cage and drum.

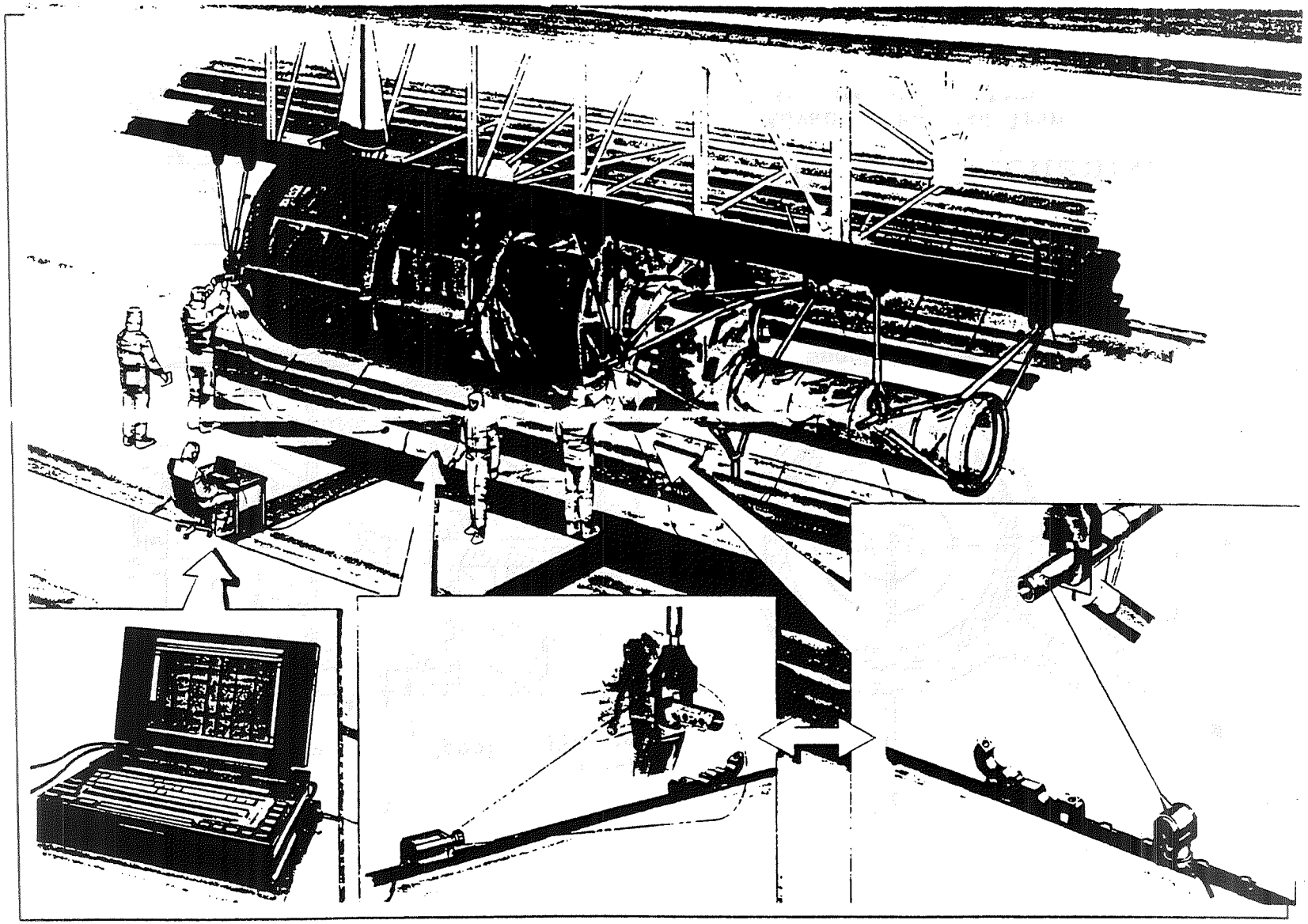


Figure1-1: APT-MS Work Environment Senario

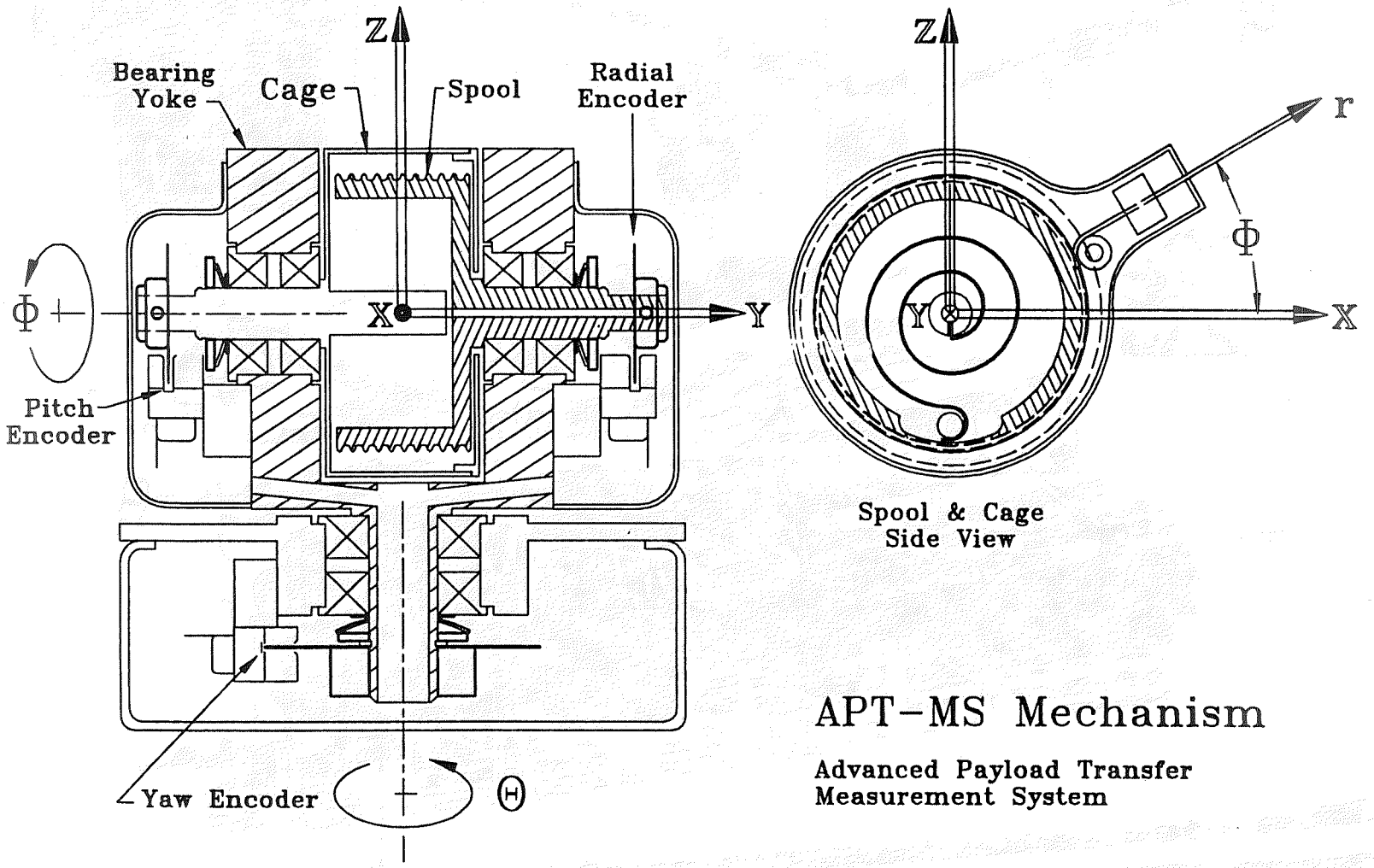


Figure1-2: Assembly Schematic

II

ERROR SOURCES

Estimatable sources of mechanical error considered, fall into two categories:
Cable clearance errors, Bearing friction errors.

2.1 CABLE GUIDE

In this case the term "cable guide" is an inverse of its usual meaning. Cable tension is used to guide cage rotation about both horizontal and vertical axes and thus obtain pitch and yaw angle readings. To understand this function, consider the diagram in Figure 2-1.a. Initially the cable is horizontal and the pitch angle reads zero. Then the object to which the right end of the cable is attached moves upward. This should be reflected with a change in pitch angle reading which is controlled by cage rotation. Only tangential components (i.e. perpendicular to the radial) of cable tension are able to produce moments about cage (or yoke) rotational axes and thus bring the cage cable guide in line with a new cable direction. The two error sources indicated above prevent a perfect alignment.

2.2 CABLE CLEARANCE

A cable guide contains the last 2 points where cable contacts the cage. Clearance should be minimized at these points. To whatever extent the cable may move without contacting the last bearing point, an error is introduced. Figure 2-1.b shows orifices of radius R being guided by a cable of a smaller radius r . Initially the cable is taken as horizontal where the object to which it is attached has been moving down from above the horizontal axis. The cable is therefore bearing on the lower surfaces of guide orifice peripheries. If the moving object then reverses its direction and returns upward, a distance $e = 2(R-r)$ is traversed by the cable between lower and upper exit contact points before a moment producing force can be applied to rotate the cage. This rotational lag is cable clearance error. It is approximated as Er for small angles ϵ as indicated in Figure 2-1.b .

2.3 BEARING FRICTION

As cable presses against the cable guide exit orifice, shown in Figure 2-1.a, resistance to cage rotation is expressed as a moment M, causing the cable to deflect until a tangential component of cable tension produces sufficient counter moment for equilibrium. Cable deflection results in a pointing error Er. Similar deflections are caused by rotation resistance about the vertical bearing axis.

An error model attempts to capture as many of the contributing effects as possible. Information from several bearing vendors contributed to the model. Computations reflect those methods that consider the somewhat more complicated and variable static loadings and geometry of this bearing application. Computations using MathCad 5.0 are displayed on the following pages:

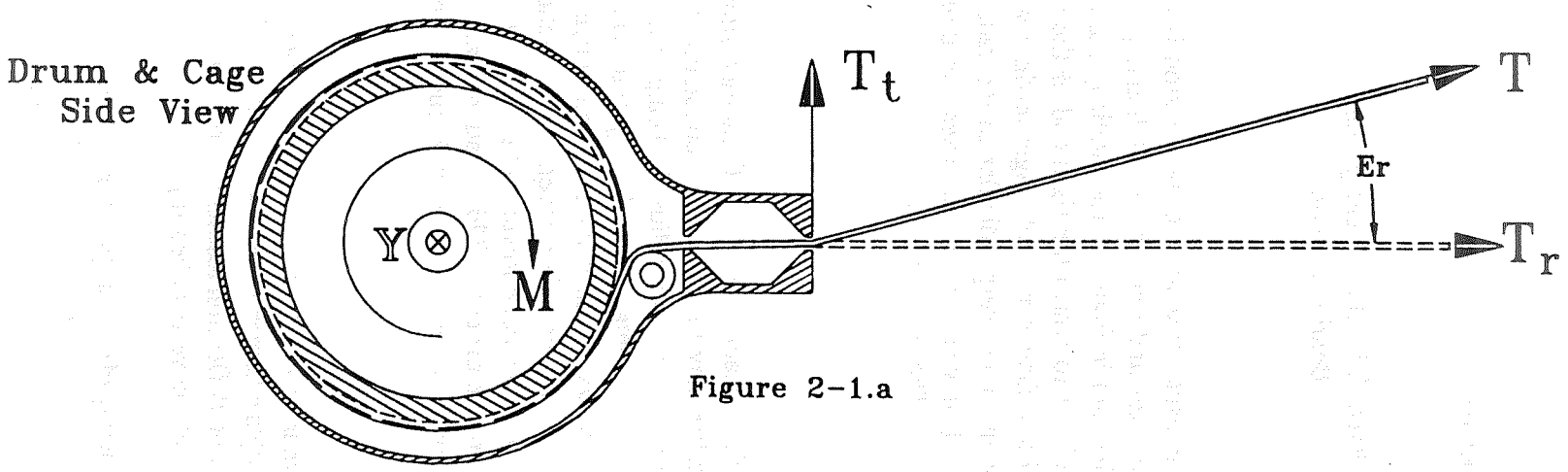


Figure 2-1.a

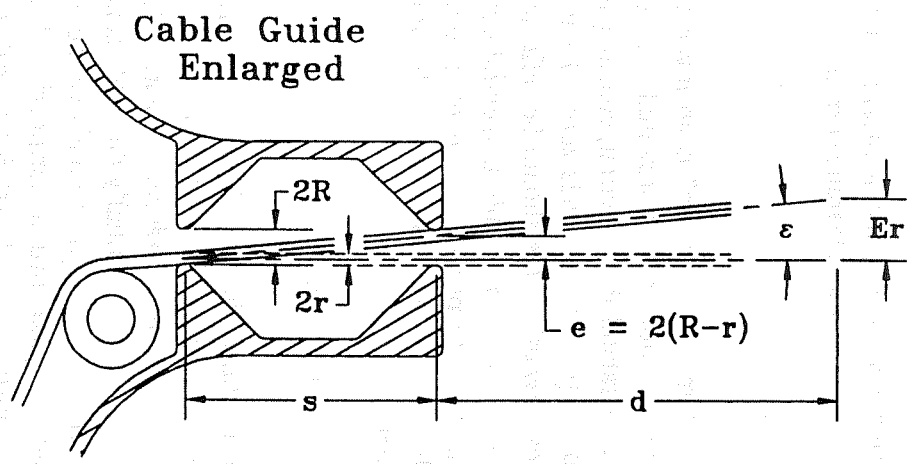


Figure 2-1.b

$$\tan(\varepsilon) = \frac{e}{s} = \frac{Er}{d + s}$$

$$Er = \frac{e}{s}(d + s)$$

$$Er = e\left(\frac{d}{s} + 1\right)$$

Figure 2-1: Error Sources

APT - MS

ERRORS from Rotational Resistance

Tension in cable
$T := 5.000$

Spring force, z component
$Sz := 0$ (neglect)

$d := 24.0$
from tube exit to cable end

$rr := 1.0$ reel radius
$l := 2.0$ from y axis to tube exit
$a := 1.5$ from y axis to cable guide
$b := 0.0$ y distance free cable from reel to guide

$\alpha := \operatorname{asin}\left(\frac{rr}{l}\right)$ $\sin\delta := \frac{b}{\sqrt{a^2 + b^2 - rr^2}}$

$i := 1..2$ z, y axis bearings

$j := 1..4$ ball locations

$\gamma := 90\cdot deg$ z axis from vertical

$\theta := 90\cdot deg$ x axis from horizontal

$\phi_i :=$ pitch angle

$45\cdot deg$
$0\cdot deg$

always zero

BEARING GEOMETRY

$\beta_i :=$ bearing contact angle

$10\cdot deg$
$11\cdot deg$

$c_i :=$ center distance of bearing set measured

.4155
.3798

$pd_i :=$ pitch diameter of bearing measured

.800
.632

effective center distance
$C_i := c_i + pd_i\cdot\tan(\beta_i)$ $\rho_i := \frac{C_i}{2}\cdot\cos(\beta_i)$

effective pitch diameter
$PD_i := pd_i + c_i\cdot\cot(\beta_i)$ $\rho_i := \frac{PD_i}{2}\cdot\sin(\beta_i)$

ρ_i

0.274
0.247

effective moment arm

$\beta c_i := \operatorname{atan}\left(\frac{C_i}{PD_i}\right)$

ρ_i

0.274
0.247

check

$\frac{\beta c_i}{deg}$ check

10
11

BEARING LOADS

assumes all loads additive

Sources:

Weight, preLoad, Tension in cable

$W_i :=$

5.000
1.000

$pL_i :=$

0.0
0.0

$Tr_1 := T\cdot\cos(\phi_1)$

$Rx := T\cdot(2 - \cos(\alpha))$ $Rx = 5.67$

$Rz := T\cdot\sin(\alpha) + Sz$ $Rz = 2.5$

$Tr_2 := \sqrt{Rx^2 + Rz^2}$

$Wr_1 := W_1\cdot\sin(\gamma)$

$Wr_2 := W_2\cdot(\sin(\theta) + \cos(\theta)\cdot\cos(\gamma))$

$Ta_1 := T\cdot\sin(\phi_1)$

$Ta_2 := T\cdot\sin\delta$

$Wa_1 := W_1\cdot\cos(\gamma)$

$Wa_2 := W_2\cdot\cos(\theta)\cdot\sin(\gamma)$

moment arms:

$r_1 := 0$ always zero

$r_2 := 0$ Tsinδ not zero but neglected for simplicity

$a_i :=$

1.837
.9283

Radial Loads: $Pr_i := Wr_i + Tr_i$

Pr_i

8.536
7.197

Axial Loads: $Pa_i := Wa_i + Ta_i + pL_i$

Pa_i

3.536
0

Moments: $M_i := r_i\cdot Pa_i + a_i\cdot Pr_i$

M_i

15.68
6.681

APT - MS

ERRORS

from

Rotational Resistance

BEARING REACTIONS

$$R_{i,3} := \frac{Pr_i}{2 \cdot \cos(\beta_i)} \qquad R_{i,2} := \frac{Pa_i}{2 \cdot \sin(\beta_i)} \qquad R_{i,1} := \frac{M_i}{2 \cdot \rho_i} \qquad R_{i,4} := R_{i,1} + R_{i,2} + R_{i,3}$$

$R_{i,3}$

4.334
3.666

$R_{i,2}$

10.18
$-1.003 \cdot 10^{-15}$

$R_{i,1}$

28.607
13.54

$R_{i,4}$

43.121
17.205

ROTATION RESISTANCE

Rolling Resistance of Bearing Ball

f := .0018 (this factor obtained by R.B. from F.A.G. bearing Co.)

$$Mrr_i := \frac{pd_i}{2} \cdot f \cdot \left(2 \cdot \sum_{j=1}^{4} R_{i,j} \right)$$

Mrr_i Moment rotation resistance

0.124
0.039

ERROR from ROLLING RESISTANCE

$$Er_i := \left(\frac{Mrr_i}{T \cdot 1 \cdot \cos(\phi_i)} \right) \cdot d$$

Er_i Error resulting from Mrr

0.422
0.094

2.4 ERROR MODEL

An explanation of model parameters will serve to both indicate error sources and describe APT device functions (also see Figure 1-2). Symbols are listed in roughly the order of their introduction in the error model on the previous two pages, together with a discussion of their significance where appropriate:

T cable Tension rotates the cage (beneficial) and increases bearing loads (detrimental).

Sz radial Spring force is included for completeness but neglected in calculation.

d distance from cable guide exit to the outer end of the cable. Two feet or less is believed to be the range of interest for most payload transfers.

rr reel (or drum) radius used to compute the radial component of spherical coordinates.

l length from y rotational axis to cable guide exit.

a distance from y rotation axis to cable guide entrance.

b y distance spanning free cable between leaving drum to contact with cable guide.

α angle wrt x-axis of internal cable between drum and cage cable guide.

δ angle between internal cable tension component and the y-axis.

i $i = 1$ pertains to bearings on the y-axis, $i = 2$ those of the z-axis.

j $j = 1$ to 4 pertains to bearing ball locations in each bearing set (cage or yoke).

The foregoing assumes for purposes of explanation that, the z-axis is vertical, x and y axes are horizontal, and the cable is in an x-z plane. The error model allows the z-axis to tilt away from the vertical causing the x-y plane to be tipped wrt the horizontal. Bearing loads attributed to component weights are affected.

γ angle between z-axis and the vertical.

θ angle of x-axis from intersection of x-y plane and horizontal plane.

ϕ pitch angle of cable wrt x-y plane.

Pitch angle must be limited since the error in rotational position measurement about the z-axis becomes unbounded as the pitch approaches 90degrees.

2.4.1 BEARING GEOMETRY. Rolling resistance is a function of normal forces on the bearing balls which in turn depend on bearing geometry: bearing pitch diameter (pd); center distance between bearings (c); contact angle between ball and race (β). Effective moment arm for eccentric bearing loads is a computed quantity depending on the above physical parameters. Effective center distance (C) and pitch diameter (PD) are intermediate computations on the way to finding the effective moment arm (ρ).

2.4.2 BEARING LOADS. Contributions to aggregate bearing loads come from three major sources: weight of mechanical components (W); bearing preloads (pL); and tension in the cable (T). These have been broken into their contributions to orthogonal components, (Pr) radial or perpendicular to bearing axes, (Pa) axial or parallel to bearing axes, and (M) the aggregate moment that both of these forces cause. Of the load sources, T and W are treated in some detail. Preload has merely been added to the axial component of aggregated bearing loads. Physically, preload behavior is somewhat more complicated. In this application preloads are expected to be small therefore modeling of them has been postponed. They are included to recognize there presence but then neglected.

2.4.3 BEARING REACTIONS are the normal forces (R) acting on a diameter of the bearing ball where it is in contact with the races. Reactions are modelled as if they are all applied to just 4 bearing balls located in the plane of the aggregated loads (Pa and Pr). Two balls for each bearing are at the extremes of their respective pitch diameters.

2.4.4 ROLLING RESISTANCE is expressed as a moment (Mrr) resulting from tangential forces in the same fashion as friction on a sleeve bearing. The factor (f) is used like a friction coefficient but is not the result of sliding contact. We do not have a very certain number for this crucial quantity.

2.4.5 ERROR (Er) varies directly with rolling resistance (Mrr) and cable played out (d) and inversely with cable tension (T), distance from rotation axis to cable exit (l), and cos(ϕ) of pitch angle. Since Mrr is also a function of T, the expression can be put in a form similar to

$$Er(T) = K + C * (W/T) \qquad \text{where K and C are constants}$$

Increasing T reduces the error attributed to component weights (W) but does not affect those caused by T, expressed here as K.

Clearly, the cable direction cannot be allowed to approach the z-axis since a pitch angle of 90 degrees results in unbounded error.

Only compactness of the mechanism prevents the obvious alternative for error mitigation: increasing the distance from rotation axis to cable exit (l).

Rolling resistance can be reduced in various ways: reduce factor f ; increase bearing separation c; reduce component weights; increase cable tension; etc.

III

DESIGN DETAILS

3.1 DRUM & CAGE

Subassembly shown in Figure 3-1 calls out the various components that were not included in schematic Figures 1-2 and 2-1. Major components listed first are called out in the central column. Most components are indicated in both top (left) and side (right) views. Parts will be referenced by their number in the discussion that follows.

3.2 FUNCTION

Cable 19 is drawn from a helical storage groove in drum 3 . Tension is maintained in the cable with retraction "clock" spring 4 . The spring is anchored to the cage 2 by insertion of a tab at the inner end into a slot in post 5 , and fixed to drum 3 by a 180 degree loop at the outer end engaging the spring stud 10 . Cable is guided to the exit with rollers 6 and finally with the cable guide 1 . Set screw 11 fastens cable to drum. Counter weights 12 mounted on stud 13 serve to balance the heavier exit end of the cage subassembly on its bearing shaft.

3.3 CABLE GUIDE

Zero Clearance Cable Guide, part no.1, Figure 3-2, is a response to cable clearance error discussed in section 2.2 and illustrated in Figure 2-1 above. Thread on the HOUSING left end attaches to cage 2 with a nut. Thread on the right end is used by a CAP to compress the CORE. The CORE has two greatly reduced sections which act as thin-walled tubes. Axial compression causes the tubes to bulge radially inward reducing cable clearance which can be adjusted to the minimum tolerable. To insure that tubes bulge inward instead of outward as is their wont, voids surrounding the tubes are filled with silicon rubber. Rubber is considered incompressible under normal conditions. As the volume containing it is compressed, the metal housing prevents outward expansion therefore it will move against the tubes forcing them into the cable path. The core and its tubes are teflon to reduce cable friction.

3.4 CAGE

Guide cage 2 and its components are detailed in Figure 3-3. Rollers 6 are identical and made of teflon to reduce friction. Likewise the studs on which they are mounted are identical. Roller studs are mounted on the cage in " T " slots so their position can be adjusted to accommodate a range of cable and roller diameters. Thread on the snap ring groove end of the studs serves only to secure the outer (radial direction) roller cavity dust cover and has no function in the inner roller location.

APT-MS
Drum & Cage Sub-Assembly

no.	description	req'd.	mat'l.
1	cable guide	1	sub-assmb
2	guide cage	1	aluminum
3	cable drum	1	aluminum
4	clock spring	1	s.s.
5	spring post	1	aluminum
6	guide roller	3	teflon
7	roller stud	3	c.r.s

no.	description	req'd.	mat'l.
8	T-nut, #5-40	3	c.r.s.
9	snap ring	3	steel
10	spring stud	1	c.r.s.
11	set screw	1	#5-40
12	contr. wt.	N	Pb
13	c. w. stud	1	c.r.s.
14	std. nut	3	#5-40

no.	description	req'd.	mat'l.
15	std. nut	1	.25-20
16	drum cover	1	aluminum
17	roller cover	1	aluminum
18	flt. hd. scrw.	4	#2-56
19	cable, .030"d.	1	coated
20	cable terminal	1	sub-assmb
21	spring cover	1	aluminum

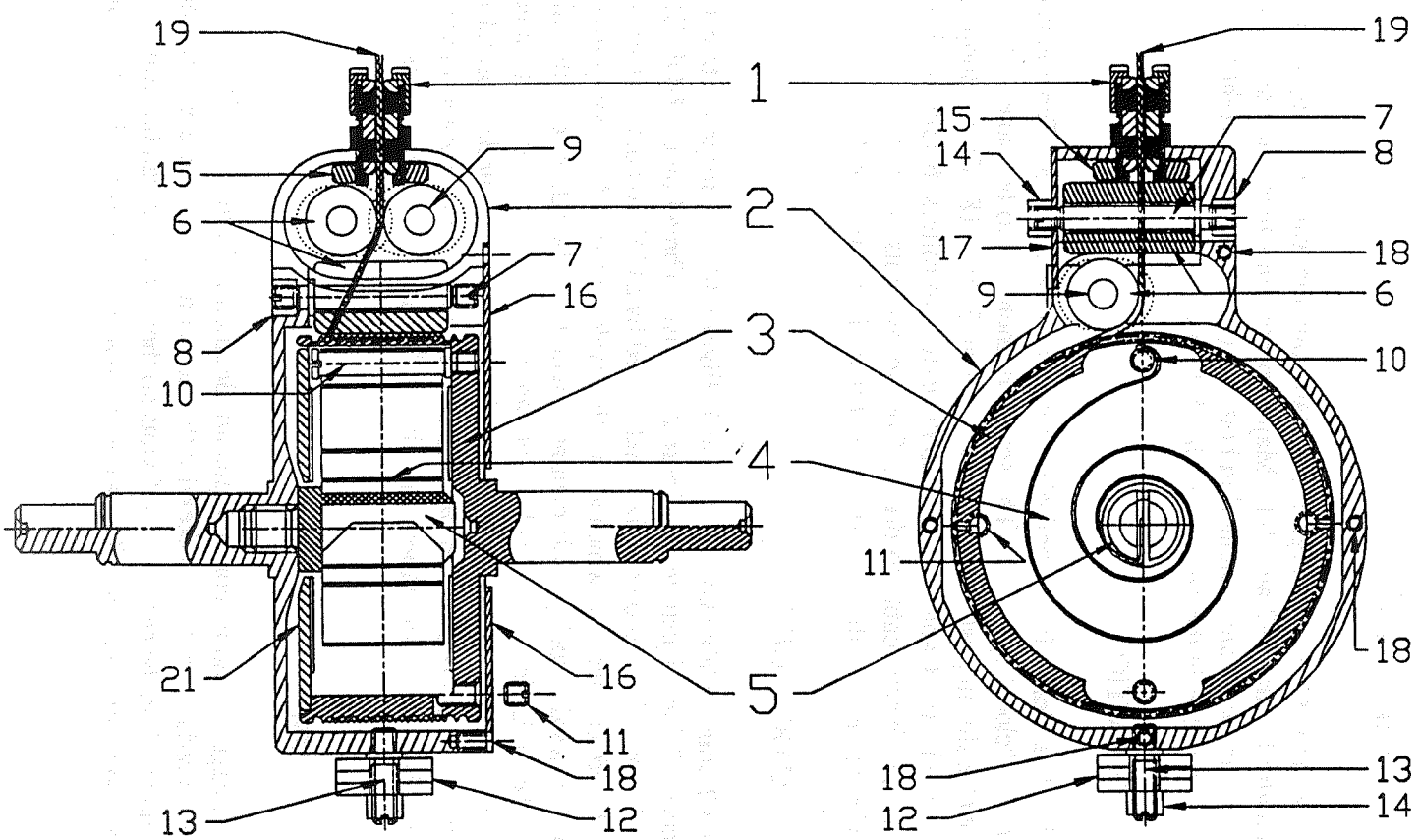


Figure 3-1: Drum & Cage Sub-Assembly

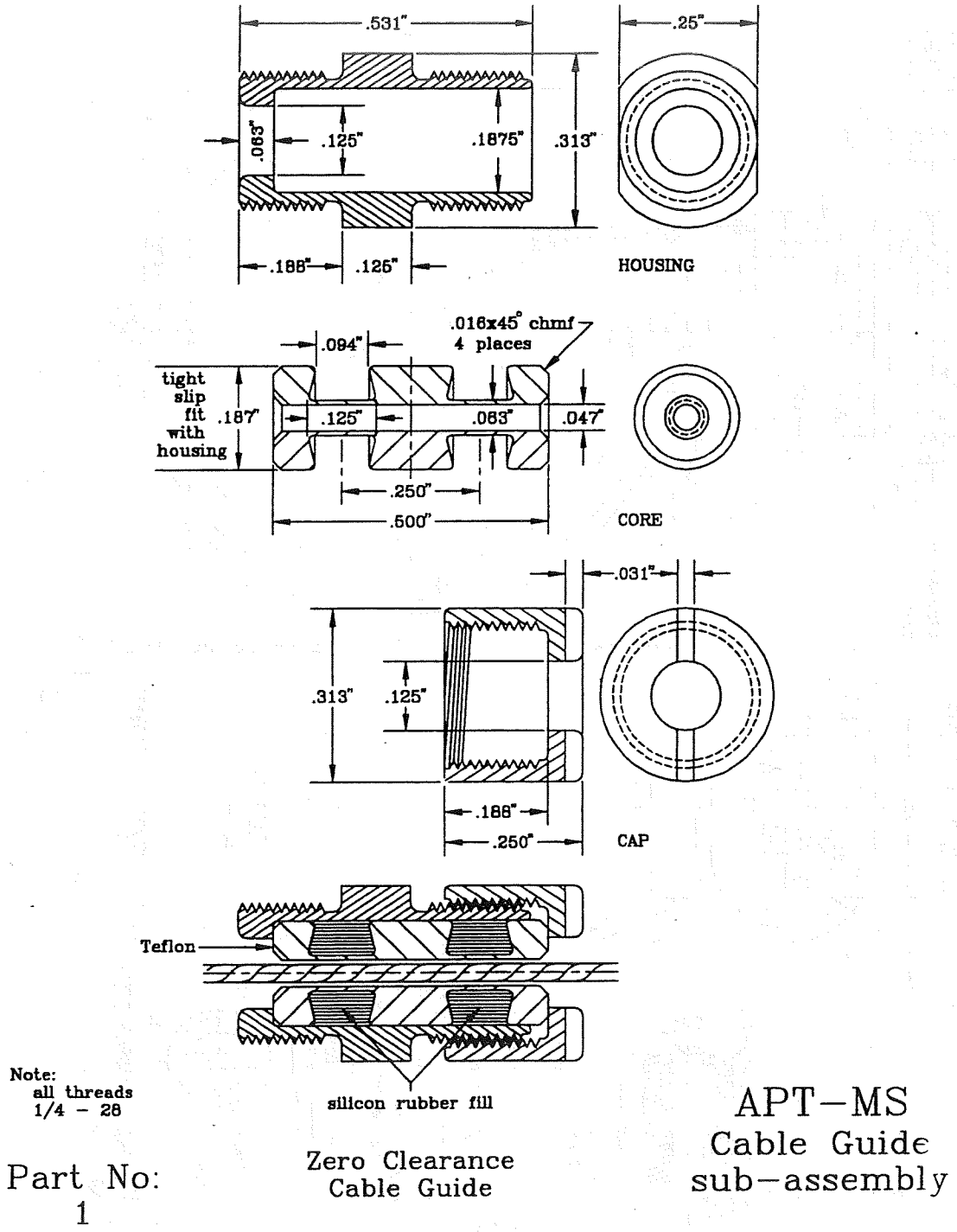


Figure 3-2: Zero Clearance Cable Guide

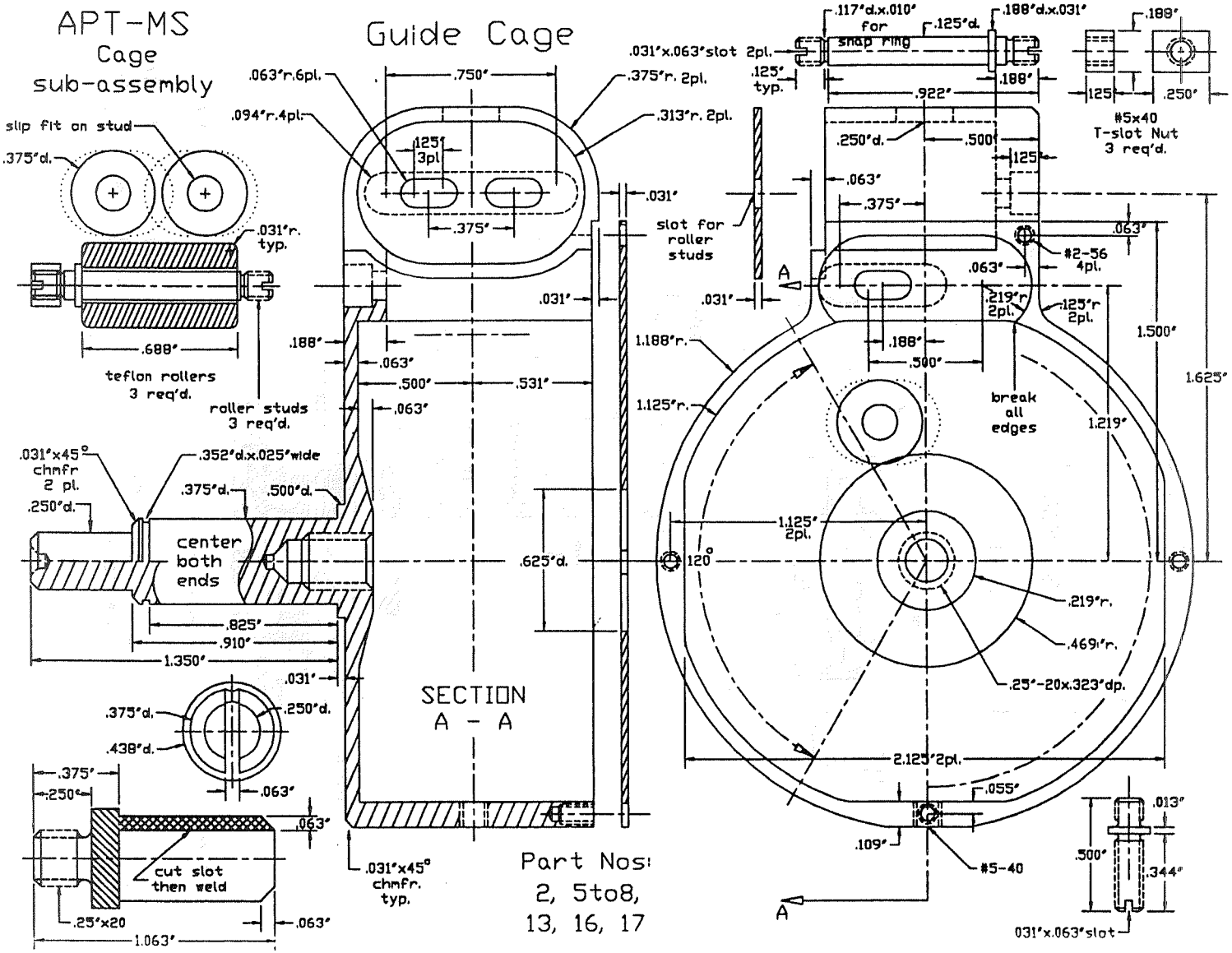


Figure 3-3: Cage Details

3.5 DRUM

Cable drum 3 and components are detailed in Figure 3-4. Explanation of the crater like feature opposite the left end of the shaft may be appropriate. When Drum 3 and Cage 2 are being assembled, the spring 4 is inaccessible and invisible. The crater fits over the right end of spring post 5 forcing the spring against the base of the post at its left end thus centering it in the spring cavity.

3.6 TERMINALS

Cable terminals, Figures 3-5 and 3-6, part no. 20, are not shown in the DRUM & CAGE subassembly. Friction error caused by these features has not been evaluated.

3.6.1 RING - ENDS, Figure 3-5, allow 360 degrees of pitch rotation but only 90 degrees of yaw motion. Stand-off may be required to prevent interference with the yaw motion Roll rotation is not required and may be detrimental to performance. A "fish line swivel" can be introduced between terminal and cable to nullify some of the latent cable roll which may be present due to varying tension.

Reference points on an object being tracked are not completely fixed relative to that object when ring type terminals are used. If the center of the least diameter on the mounting finial is taken as reference then it is fixed for pitch rotation of the terminal. Yaw motions cause this reference to move thru an arc:

$$s = r\theta = +/- (.086")(45\pi/180) = +/- .068"$$

where r = .086" is the least finial radius and yaw is +/-- 45deg.. If the center of yaw rotation is the reference then it is fixed relative to yaw rotations but moves with pitch rotation. The same formula above describes the arc length of travel.

3.6.2 BALL - ENDS, Figure 3-6, achieve a stable reference point. Disadvantages to this approach include: higher cost, less compact, and rotations limited to a cone angle of less than 180 degrees. Practical cone angles are much less. Angles of less than 90 degrees are used in the design shown in Figure 3-6. Friction increases with cone angle of the finial due to higher normal forces caused by "wedging" action of the ball in its socket.

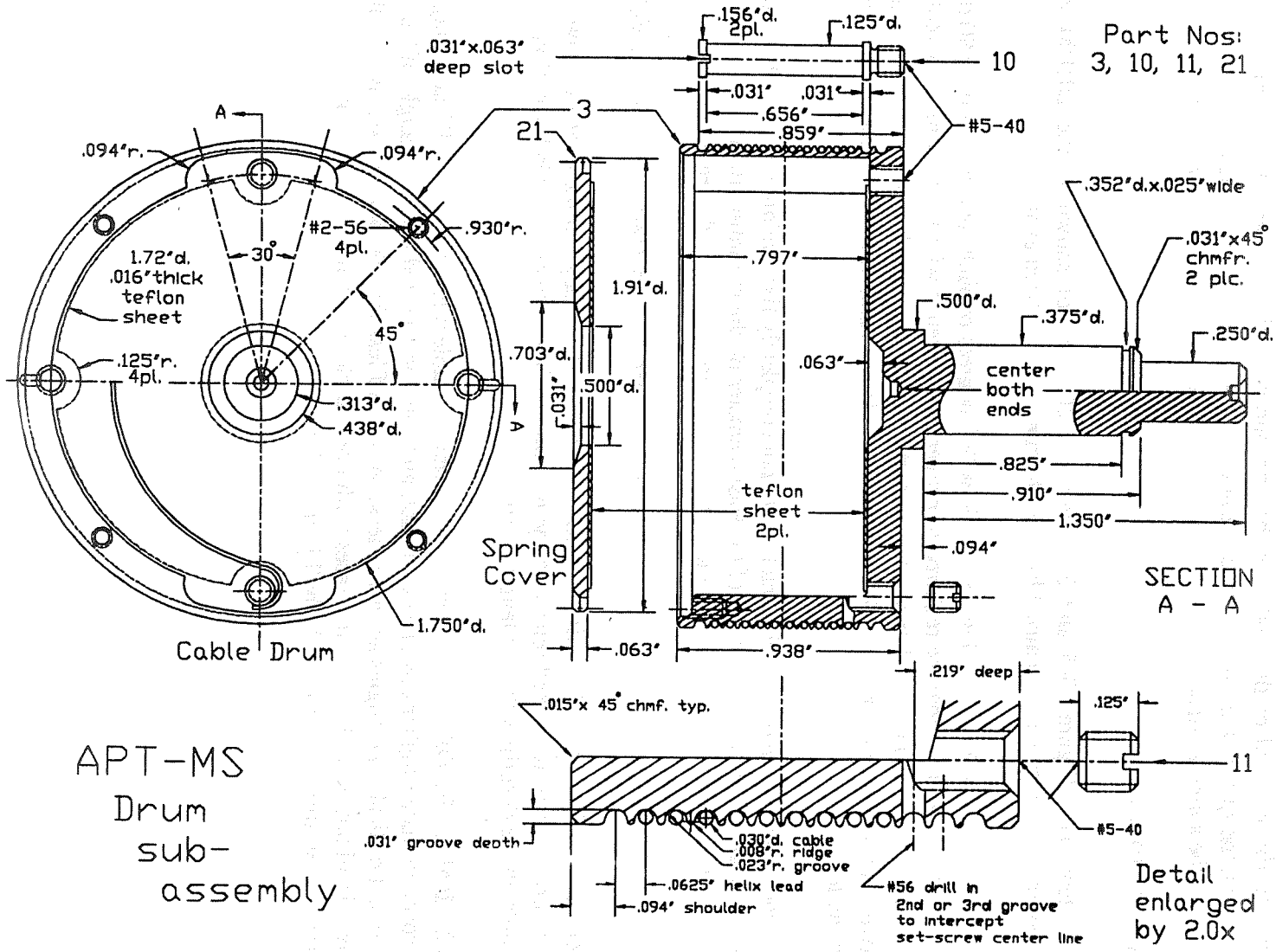


Figure 3-4: Drum Details

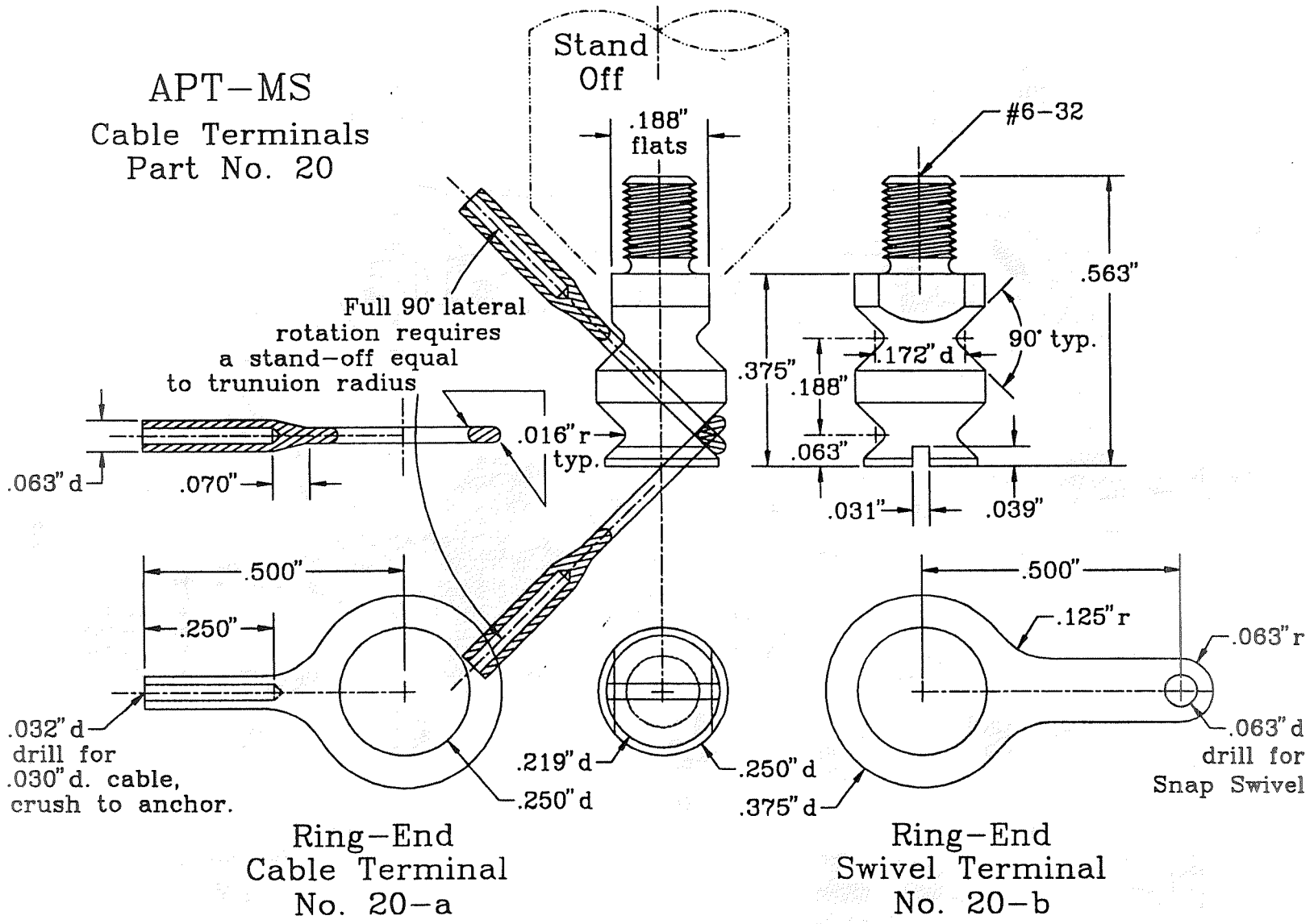
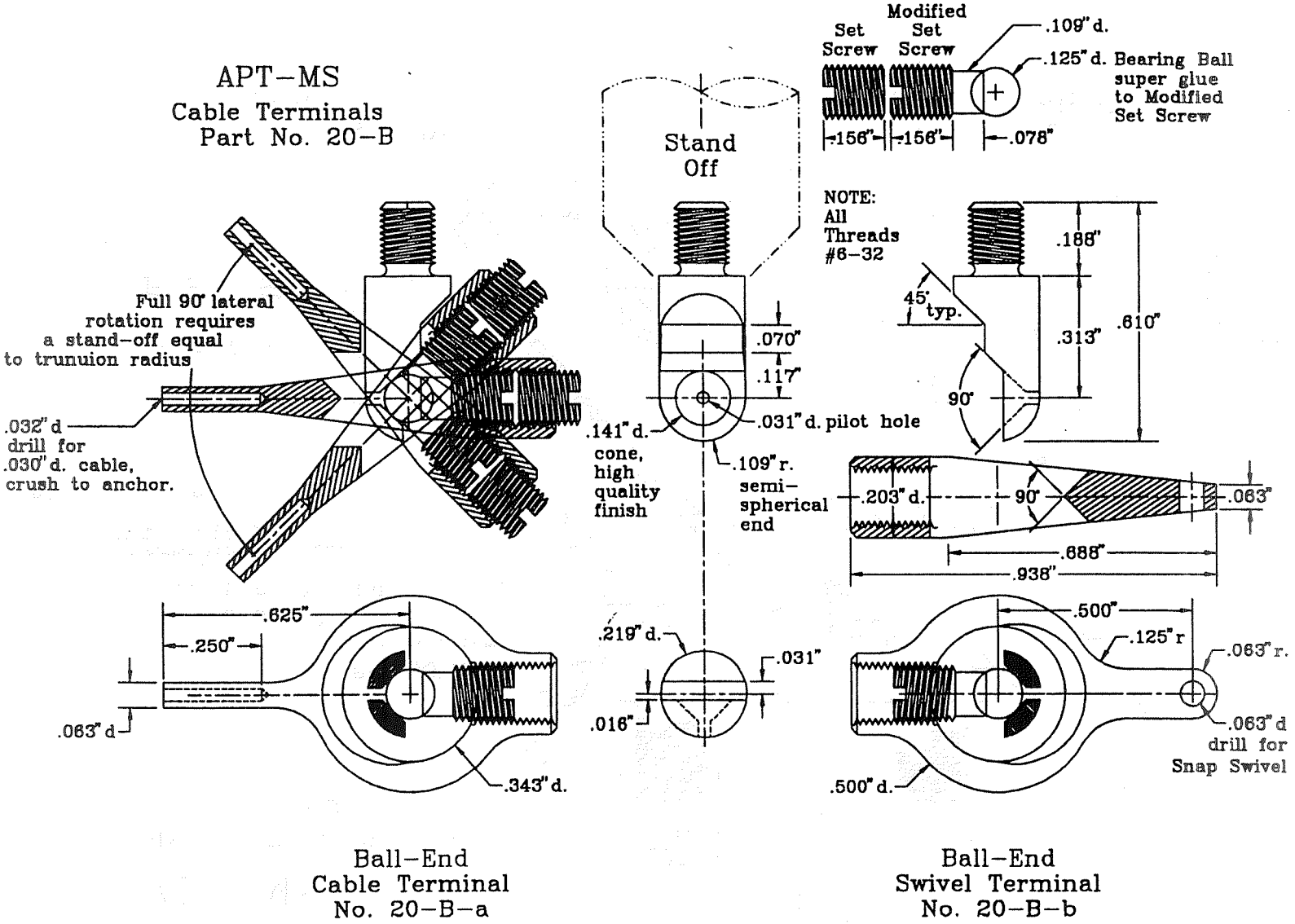


Figure 3–5: Ring–End Cable Terminals



Ball–End
Cable Terminal
No. 20–B–a

Ball–End
Swivel Terminal
No. 20–B–b

Figure 3–6: Ball–End Cable Terminals

IV

DISCUSSION

4.1 MEASUREMENT ERROR

4.1.1 GENERALITIES. Repeatability is more important than absolute positional accuracy for most of the tasks envisioned for the APT mechanism. This observation does not affect the course of development for the device since most identifiable sources of non-repeatability are the same as those causing measurement error.

Goals for positional accuracy are on the order of .050" at a distance of 24" from the device. "Worst case scenario" vector error was estimated .800" at the same distance for the design in-progress when the error model was introduced. An order of magnitude difference is not encouraging.

Error models that accurately reflect loading and geometry of the operational environment are most apt to be useful. Effort in this direction produced the model described in section 2.3.

Uncertainties regarding model input parameters, rolling resistance factor and weights in particular, leave the above error estimates in some doubt.

Usefulness of the model will come after the hardware is realized. It can then be calibrated and validated or revised as a predictive tool toward further development. Until then it serves to provide an understanding of the influence of relevant parameters.

4.1.2 CABLE CLEARANCE. Zero clearance would eliminate this source of guide error. In the real world zero clearance is unlikely to be achieved without encountering another Achilles heel in the mechanism: friction. After reducing sliding friction as much as possible and clearance as much as is tolerable (with the device proposed in section 3.3 above or some other scheme with the same goals), error can be further diminished by increasing distance between the last contact points of cable to guide (Figure 2-1.b).

4.1.3 BEARING FRICTION. Using the model described in section 2.4 as a guide, it appears that the effect of moment on bearing loads and therefore rolling resistance is likely to be greater than that of the forces causing them, for this mechanism design. Reducing the ratio of a/ρ would be beneficial (where "a" is the moment arm of eccentric radial loads and ρ is the effect of space between bearings in a pair). Smaller bearing pitch diameters and rolling resistance coefficient are bearing parameters that would help. Reducing component weights helps. Increasing tension reduces the effect of component weights. Tension appears to be neutral in effect, with rolling resistance being offset by greater turning force at the cable guide. Increasing the length " l " from bearing axis to cable guide exit is an easy way to improve performance.

4.1.4 EXCLUSION CONES. Unbounded errors result from operating the mechanism at pitch angles ϕ approaching 90 degrees. An operational range must be established within acceptable limits of error. Unacceptable regions can be visualized as cones of revolution about the z-axes. APT device cannot approach $\phi = -90$ degrees because of interference with the mechanism base. The current prototype permits operation at values of $\phi = +90$ degrees so that exclusion cone errors can be explored with the hardware. Operational units would have to be restricted either physically or electronically with appropriate warnings given to the user. In the scenario of Figure 1-1, where motions are always near a vertical plane, more accuracy may be achieved by mounting the device with the x-y plane parallel to the plane of motion rather than horizontally as shown.

4.2 DESIGN ALTERNATIVES

4.2.1 OPTICAL DEVICES usually have better performance and cost more than mechanical counterparts. An inexpensive optical alternative is being pursued in parallel with the mechanical prototype. Optical solutions are beyond the scope of this discussion of a mechanical option.

4.2.2 METAL TAPE. Replacing the guide cable in the mechanical prototype with a metal tape like that found in tape measure rulers has the advantage of stiffness [1]. Since metal tape with a curved cross-section resists bending about both cross-section axes it can deliver a tangential force to guide its cage with negligible tape defection. Better accuracy can be expected on this account. Placing the major bending axis in the x-z plane reduces exclusion cones for any give level of accuracy. Weight / length of tape is greater than that of the cable it replaces therefore errors due to sagging may increase also. These would doubtless be negligible at operational distances of 24 inches.

4.2.3 AIR BEARINGS. Greatly reduced friction obtained would come at a cost. This possibility was considered but rejected without research because of a belief that cost and time to develop it would exceed constraints of an inexpensive, easily portable measurement device.

4.2.4 SHAKE DOWN. Introduction of a vibrator to brake loose stick-slip friction may enhance performance. Frequencies and amplitudes must be compatible with the optical encoders.

4.3 EVALUATION

4.3.1 ACCURACY. Comparison with point to point readings of a digitally controlled milling machine, lathe, or robot to which the cable is attached could evaluate accuracy. Increments small enough that no response is detected in any direction can then be increased until a reading is evident. This may be done in the direction of each spherical coordinate of the mechanism to identify likely sources of error. Permutations and combinations of coordinates can also be tested. Orientation of all device axes should be varied to evaluate the effect of component weights.

4.3.2 REPEATABILITY. Tests repeated over large displacements of each spherical coordinate should be used following the pattern described for accuracy testing.

V

CONCLUSIONS

5.1 SUMMARY

- Error Model is based on mechanism attributes as far as that information is available, including: geometry, component weights, cable tension, device orientation, bearing characteristics, etc.
- Error indicated from worst case scenario is an order of magnitude greater than desired.
- Detailed Design of second prototype, produced in 2 versions depending on spring width, has some parametric flexibility where contact with the cable is concerned.
- Hardware is being fabricated for the narrower of two spring widths considered.

5.2 CONCLUSIONS

- Error model results are inconclusive because of input data uncertainties.
- Calibration of the error model requires hardware test data.
- Evaluation rather than prediction is the best use of the error model.

5.3 RECOMMENDATIONS

- Build and test the second prototype.
- Calibrate the error model.
- Use the error model to identify the greatest error sources.
- Consider design alternatives: replacement of the cable with metal tape in particular.

REFERENCES

[1] Malladi, N.S., "Rack Insertion End Effector (RIEE) Guidance", 1994 Research Reports, NASA / ASEE, Summer Faculty Fellowship Program, NASA CR-197448, p.339.

[2] Lopez del Castillo, E, Advanced Payload Transfer Measurement System 30% Design Review, DM-ASD, NASA / KSC, FL 32899, (unpublished document), 2/21/95.

[3] Lopez del Castillo, E, Advanced Payload Transfer Measurement System 60% Design Review, DM-ASD, NASA / KSC, FL 32899, (unpublished document), 4/20/95.

1995 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

S3-47
7743
p-36

A STUDY OF VARIOUS METHODS FOR CALCULATING LOCATIONS OF LIGHTNING EVENTS

Dr. John R. Cannon
Professor and Chair
Mathematics Department
University of Central Florida
Orlando, Florida

KSC Colleague - Carl L. Lennon
Atmospheric Science

Contract Number NASA-NGT-60002
Supplement 19

August 22, 1995

Acknowledgments

The author wishes to thank Carl Lennon, Lanna Maier, John Madura, Ray Hosler, Kari Stiles and the management of the NASA's Kennedy Space Center for the opportunity to spend an enjoyable summer conducting research on the LDAR equations. Also, the author wishes to thank William Koshak of NASA's Marshall Space Flight Center for several interesting private communications. Thank you and best wishes to all.

Abstract

This article reports on the results of numerical experiments with various techniques such as linear least squares, nonlinear least squares, statistical estimates, cluster analysis and angular filters for calculating the location of lightning events from time of arrival data for the Lightning Detection and Ranging (LDAR) System at Kennedy Space Center.

Summary

This article reports on the results of numerical experiments on finding the location of lightning events using different numerical methods. The methods include linear least squares, nonlinear least squares, statistical estimations, cluster analysis and angular filters and combinations of such techniques. The experiments involved investigations of methods for excluding fake solutions which are solutions that appear to be reasonable but are in fact several kilometers distant from the actual location. Some of the conclusions derived from the study are that bad data produces fakes, that no fool-proof method of excluding fakes was found, that a short base-line interferometer under development at Kennedy Space Center to measure the direction cosines of an event shows promise as a filter for excluding fakes. The experiments generated a number of open questions, some of which are discussed at the end of the report.

Table of Contents

I. Introduction

The Lightning Detection and Ranging (LDAR) System developed over a number of years by Carl Lennon and colleagues at the Kennedy Space Center is based on the times of arrival at seven locations of electromagnetic radiation in the 66MH frequency range emitted by electric charge movement in thunder storms. The differences in the times of arrivals are converted into differences in distances from the point of origin (x, y, z) of the radiation to the various receiving sites $\#i = 0, \cdots, 6$ located at $(x_i, y_i, 0)$, $i = 0, \cdots, 6$, where site #0 is taken to be the origin $(0, 0, 0)$ of the local rectangular coordinate system with the positive x-axis directed east, the positive y-axis directed north and the positive z-axis directed up. The geographic locations are shown in Figure 1 along with their latitude and longitude coordinates. Relative to the local coordinate system the site locations in kilometers from Site #0 are

	x_i	y_i	z_i
Site #0	0	0	0
Site #1	3.255	9.462	0
Site #2	7.466	.014	0
Site #3	5.532	-7.056	0
Site #4	-3.854	-5.792	0
Site #5	-8.424	-1.007	0
Site #6	-3.738	7.460	0

Remark: The origin $(0, 0, 0)$ is located at the top of the antenna at Site #0 and the heights of the antennas at the other sites have been adjusted taking the curvature of the earth into account to be nearly in the $xy-$plane. A recent survey has shown the various antenna to be within a meter or so of the $xy-$plane.

Denoting the time of arrival of the signal at Site $\#i$ by t_i, $i = 0, 1, \cdots, 6$, and the distance from the point of origin of the signal (x, y, z) to $(x_i, y_i, 0)$ (Site $\#i$)by

$$d_i = \sqrt{(x - x_i)^2 + (y - y_i)^2 + z^2}, \tag{1}$$

we see that

$$d_i - d_j = u_{ij}, \qquad i, j = 0, \cdots, 6, i \neq j, \tag{2}$$

where

(3) $$u_{ij} = c\left(t_i - t_j\right)$$

and

(4) $$c = 299792.458 \text{ km/sec}$$

is the measured speed of light. Current practice is to use Site #0 as the trigger for the initiation of a time of arrival measurement at all sites. Transmission times to and from the other sites are taken into account in the determination of the $t_i, i = 1, \cdots, 6$. From these time measurements the six equations

(5) $$d_i - d_0 = u_{i0}, \qquad i = 1, \cdots, 6$$

which are solved for x, y and z. The equations appear to be nonlinear at first glance. However, setting

(6) $$r = d_0 = \sqrt{x^2 + y^2 + z^2}$$

we find that

(7) $$d_i = \sqrt{r^2 + r_i^2 - 2x_i x - 2y_i y}$$

where

(8) $$r_i^2 = x_i^2 + y_i^2.$$

Shifting d_0 to the right side of (5), squaring both sides, and employing (6), (7) and (8), we find after some elementary calculations that (x, y, z) satisfies

(9) $$x_i\, x + y_i\, y + u_{i0} r = v_i, \qquad i = 1, \cdots, 6$$

where z is contained implicitly in the unknown r and

$$v_i = \left(r_i^2 - u_{io}^2\right)/2, \qquad i = 1, \cdots, 6. \tag{10}$$

We note here that equations (2) for $i, j \neq 0$ do not admit this simplification and remain nonlinear in x, y and r. The pairs of three equations $i = 1, 3, 5$ and $i = 2, 4, 6$ are solved for x, y and r with z determined from

$$z = \sqrt{r^2 - x^2 - y^2} \tag{11}$$

Let $(x_{\text{odd}}, y_{\text{odd}}, z_{\text{odd}})$ and $(x_{\text{even}}, y_{\text{even}}, z_{\text{even}})$ denote the respective solutions.

If

$$\begin{aligned} |x_{\text{odd}} - x_{\text{even}}| &< \eta, \\ |y_{\text{odd}} - y_{\text{even}}| &< \eta, \\ |z_{\text{odd}} - z_{\text{even}}| &< \eta, \end{aligned} \tag{12}$$

where

$$\eta = \begin{cases} .35\text{km if } R < 7\text{km} \\ (.05)\,R\text{km if } R \geq 7\text{km} \end{cases} \tag{13}$$

and

$$R = \left(x_{\text{odd}}^2 + y_{\text{odd}}^2\right)^{1/2}, \tag{14}$$

then

$$\begin{aligned} x &= \left(x_{\text{odd}} + x_{\text{even}}\right)/2, \\ y &= \left(y_{\text{odd}} + y_{\text{even}}\right)/2, \\ z &= \left(z_{\text{odd}} + z_{\text{even}}\right)/2, \end{aligned} \tag{15}$$

On the other hand if (12) is not satisfied, then each of the 20 combinations of three independent equations from the six equations in (9) are solved and each solution is compared with the others to find at least eight solutions which satisfy the tolerance limits in (12) with R in (14) calculated

from the xy–coordinates of the solution used as the base for each comparison. Preference is given to the odd and even sites by counting them twice when they are used as the comparison base. If 8 such solutions are found to satisfy (12), then the solution (x, y, z) is determined from their average. If 8 solutions satisfying (12) cannot be found, then no solution is determined and the system proceeds to the next event. The choice of 8 stems from the combinatoric possibility of 10 good solutions if one site is down, etc. Currently, solutions are determined for 20 to 50 percent of the events observed by the LDAR system. We should note here that the even and odd combinations are nearly optimally configured to minimize the amplification of measurement errors which explains their preference in the current process. We shall expand this brief explanation in our discussion on errors in section II. To illustrate the problem the reader is referred to Figure 2 which shows the xy–plot of the ~42 solutions determined from synthetic data subjected to the addition of error samples from a uniform distribution of magnitude + or - 750 meters.

As noted in the abstract the purpose of this report is to discuss the results of numerous computer simulations which were carried out from April through August 1995. We begin with a discuss of errors in Section II and proceed to a discussion of methods considered in Section III. In Section IV, we focus on the utility of additional angular information that may become available via measurements from a short base line interferometer which is currently under development for location at the central Site #0. In Section V, we summarize conclusions and open questions for future study.

II. Measurement Errors

We begin with the system (9) which we rewrite as

$$x_i\, x + y_i\, y + u_i\, r = v_i, \qquad i = 1, \cdots, 6, \tag{16}$$

where here and in all that follows we set

$$u_i = u_{i0}. \tag{17}$$

Given any three equations $i, j, k, i \neq j \neq k$, we can write

$$A_{ijk}\, \bar{s}_{ijk} = \bar{b}_{ijk}, \tag{18}$$

where

(19) $$A_{ijk} = \begin{pmatrix} x_i & y_i & u_i \\ x_j & y_j & u_j \\ x_k & y_k & u_k \end{pmatrix},$$

(20) $$\vec{s}_{ijk} = \begin{pmatrix} x_{ijk} \\ y_{ijk} \\ r_{ijk} \end{pmatrix},$$

and

(21) $$b_{ijk} = \begin{pmatrix} v_i \\ v_j \\ v_k \end{pmatrix}.$$

Note that we have displayed the dependence of $x, y,$ and r on i, j, k via subscripts. For perfect data, the x_{ijk}, y_{ijk}, and r_{ijk} are all equal respectively. Suppose that u_i is replaced by $u_i + \varepsilon_i$ and that v_i is replaced by $v_i - u_i\varepsilon_i - (\varepsilon_i^2/2)$ in (16) above. Then, the system (18) can be written as

(22) $$A_{ijk}\ \vec{s}_{ijk}(\varepsilon_i, \varepsilon_j, \varepsilon_k) = \vec{c}_{ijk}(\varepsilon_i, \varepsilon_j, \varepsilon_k)$$

where the solution $\vec{s}_{ijk}$ denoted in (20) above has its dependence on the errors $\varepsilon_i, \varepsilon_j,$ and ε_k displayed in (22) and

(23) $$\vec{c}_{ijk} = \vec{b}_{ijk} + \vec{f}_{ijk}(\varepsilon_i, \varepsilon_j, \varepsilon_k, u_i, u_j, u_k, r),$$

where

(24) $$\vec{f}_{ijk} = -\begin{pmatrix} \varepsilon_i r + u_i\varepsilon_i + (\varepsilon_i^2/2) \\ \varepsilon_j r + u_j\varepsilon_j + (\varepsilon_j^2/2) \\ \varepsilon_k r + u_k\varepsilon_k + (\varepsilon_k^2/2) \end{pmatrix}.$$

Subtracting (18) from (22), we see that

(25) $$\vec{s}_{ijk}(\varepsilon_i, \varepsilon_j, \varepsilon_k) = \vec{s}_{ijk}(0,0,0) + A_{ijk}^{-1}\,\vec{f}_{ijk}.$$

We note two things. First, for fixed measurement errors, the overall solution error

(26) $$\vec{s}_{ijk}(\varepsilon_i, \varepsilon_j, \varepsilon_k) - \vec{s}_{ijk}(0,0,0) = O(r),$$

where the Landau Order symbol is employed here:

(27) $$g = O(r) \quad \textit{iffi} \quad |g| \le Cr$$

where C is a positive constant, i.e.,

$$x_{ijk}(\varepsilon_i, \varepsilon_j, \varepsilon_k) - x_{ijk}(0,0,0) = O\big(r_{ijk}(0,0,0)\big)$$
$$y_{ijk}(\varepsilon_i, \varepsilon_j, \varepsilon_k) - y_{ijk}(0,0,0) = O\big(r_{ijk}(0,0,0)\big)$$
$$r_{ijk}(\varepsilon_i, \varepsilon_j, \varepsilon_k) - r_{ijk}(0,0,0) = O\big(r_{ijk}(0,0,0)\big)$$

This effect of the increase in error with the distance of the event can be observed computationally. Second, the analysis of relative error in solutions of linear systems depends on a factor called the condition number of the matrix; i.e.,

(28) $$\frac{\left\|\vec{s}_{ijk}(\varepsilon_i, \varepsilon_j, \varepsilon_k) - \vec{s}_{ijk}(0,0,0)\right\|}{\left\|\vec{s}_{ijk}(0,0,0)\right\|} \le K(A_{ijk}) \frac{\left\|\vec{c}_{ijk} - \vec{b}_{ijk}\right\|}{\left\|b_{ijk}\right\|},$$

where $\|\ \|$ denotes some vector norm such as the maximum of the absolute values of the components,

(29) $$K(A_{ijk}) = \left\|A_{ijk}\right\| \left\|A_{ijk}^{-1}\right\|$$

is the condition number of the matrix A_{ijk}, and $\left\|A_{ijk}\right\|$ and $\left\|A_{ijk}^{-1}\right\|$ are the matrix norms induced by the vector norm $\|\ \|$ via the relation

(30) $$\left\|A_{ijk}\right\| = \max_{\|\vec{s}\|=1} \left\|A_{ijk}\vec{s}\right\|.$$

For derivations and discussions of vector norms, matrix norms, condition number, and relative error, see [18] or any other reasonable text on numerical analysis or numerical linear algebra. For the vector norm defined as the maximum absolute value of the components of the vector, the corresponding matrix norm is the maximum of the row sums of the absolute values of the entries in each row; i.e., for the matrix $A=(a_{ij})$

$$\|A\| = \max_i \left\{ \sum_{j=1}^{n} |a_{ij}| \right\} \tag{31}$$

is induced by the vector norm

$$\|\vec{s}\| = \max_i |s_i|, \tag{32}$$

where s_i is the i^{th} component of a vector

$$\vec{s} = \begin{pmatrix} s_1 \\ s_2 \\ \cdot \\ \cdot \\ \cdot \\ s_n \end{pmatrix}. \tag{33}$$

The point here is that the matrix A_{ijk} depends on the measurements u_i, u_j, and u_k which depend in turn upon (x, y, z). Thus, as (x, y, z) vary with fixed r, the condition number $K(A_{ijk})$ varies. See Tables 1-4. The "Y-configurations" of the odd and even numbered sites is an approximation to the optimal perfect Y-configuration which minimizes the maximum of the condition number over the $x_i, x_j, x_k, y_i, y_j, y_k, u_i, u_j$, and u_k for fixed r. Taken together the increase in error with respect to r and the variation in the condition number of A_{ijk} as (x, y, z) vary is called the Geometric Dilution of Precision (GDOP).

While GDOP is a problem for (x, y, r), it is a major problem for z. Consider the expansion of $z_{ijk}(\varepsilon_i, \varepsilon_j, \varepsilon_k)$ in its Taylor series:

$$
\begin{aligned}
z_{ijk}(\varepsilon_i,\varepsilon_j,\varepsilon_k) = z_{ijk}(0,0,0) &+ \frac{\partial z_{ijk}}{\partial \varepsilon_i}(0,0,0)\,\varepsilon_i + \\
&+ \frac{\partial z_{ijk}}{\partial \varepsilon_j}(0,0,0)\,\varepsilon_j + \frac{\partial z_{ijk}}{\partial \varepsilon_k}(0,0,0)\,\varepsilon_k + \\
&+ \frac{1}{2}\frac{\partial^2_{z_{ijk}}}{\partial \varepsilon_i^2}(0,0,0)\,\varepsilon_i^2 + \frac{1}{2}\frac{\partial^2_{z_{ijk}}}{\partial \varepsilon_j^2}(0,0,0)\,\varepsilon_j^z + \\
&+ \frac{1}{2}\frac{\partial^2_{z_{ijk}}}{\partial \varepsilon_k^2}(0,0,0)\,\varepsilon_k^2 + \frac{\partial^2_{z_{ijk}}}{\partial \varepsilon_i \partial \varepsilon_j}(0,0,0)\,\varepsilon_i\,\varepsilon_j \\
&+ \frac{\partial^2_{z_{ijk}}}{\partial \varepsilon_i \partial \varepsilon_k}(0,0,0)\,\varepsilon_i\,\varepsilon_k + \frac{\partial^2_{z_{ijk}}}{\partial \varepsilon_j \partial \varepsilon_k}(0,0,0)\,\varepsilon_j\,\varepsilon_k + \\
&+ \cdots .
\end{aligned} \tag{34}
$$

As

$$
z_{ijk}(\varepsilon_i,\varepsilon_j,\varepsilon_k) = \sqrt{r(\varepsilon_i,\varepsilon_j,\varepsilon_k)^2 - x(\varepsilon_i,\varepsilon_j,\varepsilon_k)^2 - y(\varepsilon_i,\varepsilon_j,\varepsilon_k)^2} \tag{35}
$$

it follows that

$$
\frac{\partial z_{ijk}}{\partial \varepsilon_i} = \left(r_{ijk}\frac{\partial r_{ijk}}{\partial \varepsilon_i} - x_{ijk}\frac{\partial x_{ijk}}{\partial \varepsilon_i} - y_{ijk}\frac{\partial y_{ijk}}{\partial \varepsilon_i} \right) \Big/ z_{ijk} \tag{36}
$$

and

$$
\begin{aligned}
\frac{\partial^2 z_{ijk}}{\partial \varepsilon_i\,\partial \varepsilon_j} = &-\frac{1}{z_{ijk}^3}\left(r_{ijk}\frac{\partial r_{ijk}}{\partial \varepsilon_j} - x_{ijk}\frac{\partial x_{ijk}}{\partial \varepsilon_j} - y_{ijk}\frac{\partial y_{ijk}}{\partial \varepsilon_j} \right)\bullet\left(r_{ijk}\frac{\partial r_{ijk}}{\partial \varepsilon_i} - x_{ijk}\frac{\partial x_{ijk}}{\partial \varepsilon_i} - y_{ijk}\frac{\partial y_{ijk}}{\partial \varepsilon_i} \right) + \\
&+ \frac{1}{z_{ijk}}\left(\frac{\partial r_{ijk}}{\partial \varepsilon_j}\frac{\partial r_{ijk}}{\partial \varepsilon_i} - \frac{\partial x_{ijk}}{\partial \varepsilon_j}\frac{\partial x_{ijk}}{\partial \varepsilon_i} - \frac{\partial y_{ijk}}{\partial \varepsilon_j}\frac{\partial y_{ijk}}{\partial \varepsilon_i} \right) + \\
&+ \frac{1}{z_{ijk}}\left(r_{ijk}\frac{\partial^2 r_{ijk}}{\partial \varepsilon_j\,\partial \varepsilon_i} - x_{ijk}\frac{\partial^2 x_{ijk}}{\partial \varepsilon_i\,\partial \varepsilon_j} - y_{ijk}\frac{\partial^2 y_{ijk}}{\partial \varepsilon_i\,\partial \varepsilon_j} \right) \quad , \text{ etc.}
\end{aligned} \tag{37}
$$

From spherical coordinates, we recall that $z = r\sin\phi$ where ϕ is the angle of elevation out of the xy−plane. By differentiating the equations (22) with respect to $\varepsilon_i, \varepsilon_j, \varepsilon_k$, we observe (Not Proved-although a messy induction argument looks feasible) that

(38) $$\frac{\partial^n x_{ijk}}{\partial \varepsilon_i^{\,l} \partial \varepsilon_j^{\,n-l}} = O\left(r_{ijk}^n\right)$$

and similarly for y_{ijk} and r_{ijk}. Consequently, the expansion for z_{ijk} behaves like

(39) $$\begin{aligned} z_{ijk}\left(\varepsilon_i, \varepsilon_j, \varepsilon_k\right) = z_{ijk}(0,0,0) + \\ O\Big(r_{ijk}(0,0,0)\Big[\frac{1}{\sin\phi}\left\{\varepsilon + \varepsilon^2 + \cdots\right\} + \\ \frac{1}{\sin^3\phi}\left\{\varepsilon^2 + \varepsilon^2 + \cdots\right\} + \cdots\Big]\Big), \end{aligned}$$

where

(40) $$\varepsilon = \max\left\{\left|\varepsilon_i\right|, \left|\varepsilon_j\right|, \left|\varepsilon_k\right|\right\}$$

has been employed to simplify the expression in (39). This behavior in the error in z can be seen computationally by fixing z and letting r increase. For fixed z as r increases ϕ and $\sin\phi$ tend to zero and together r and ϕ rapidly destroy the accuracy in z while one can still obtain reasonable estimations of x and y.

Before we turn to a discussion of the methods explored we shall conclude with a few comments on the data error versus the current solution method at Kennedy Space Center. First, the calculation of η in (13) is a recognition GDOP and builds an approximation of the error behavior into what the system accepts as a solution. Next, the second phase of the current solution practice is the recognition of down time of a single site for maintenance and repairs. Taken together both phases impose a strict filter on the data acceptable for a solution. As seen by the scatter of solution points in figures 2-9 and discussed above, the system of equations (16) is very sensitive to measurement error magnitude, distance of the event, and the elevation angle. Thus, it is not unreasonable to impose such a filter as bad data generates bad answers and the electronics is subject to various interference sources that corrupt various portions of the data. Given the sensitivity of the system to errors, it is feasible that an error in the data could occur which would fake an acceptable solution to the system. It would be interesting to investigate the various possible types of errors in the system which could produce such fakes and whether or not the error would be an occasional fake or a frequent one. We shall comment on this again.

III. Methods Considered

Many Combinations and permutations of the following general methods were considered:

A. Linear Least Squares for x, y and r.
B. Linear Programming for x, y and r.
C. Nonlinear Least Squares for $u_i, i = 1, \cdots, 6$.
D. Nonlinear Least Squares for x, y and r.
E. Statistical Estimation for x, y and r.
F. Cluster Analysis for x, y and r.

We shall discuss each of the above in turn. However, it is worth pointing out first that all of the above methods work well for good data. When one encounters bad data all methods considered fall apart. The study focused upon the possible mix of "good" data from four of the outer sites combined with the central site with "bad" data from two of the outer sites. Another way of looking at this would be the determination of x, y and z from the best four of the measurements $u_i, i = 1, \cdots, 6$. With these comments in mind, we turn to the discussion of the above methods.

A. Linear Least Squares. Linear least squares for all six equations in (16) can be written as

$$\mu = \min_{x,y,r} s(x, y, r), \tag{41}$$

where

$$S(x, y, r) = \sum_{i=1}^{6} (x_i\, x + y_i\, y + u_i\, r - v_i)^2 . \tag{42}$$

Linear least squares also gives the possibility of solving for an (x, y, r) using any combination of three, four, or five equations. It was observed that solutions were always generated. An example of a reasonable approach to finding a solution from the best four of the sites is to find the four sites for which the corresponding μ is the least of the 15 possible μ 's obtained from considering all combinations of four equations from the six equations in (16). The (x, y, z) resulting from that best combination could be taken as the approximation to the lightning event location. In trials with synthetic data with addition of errors selected from uniform distributions ranging up to

±.5 kilometer one could always find an error combination which would produce a fake solution, i.e., the solution corresponding to the least μ would be several kilometers away from the "correct" known synthetic solution.

B. Linear Programming. The problem of minimizing the linear function

(43) $$L(x, y, r, \eta) = \eta$$

subject to the constraints

(44) $$|x_i\, x + y_i\, y + u_i\, r - v_i| \le \eta$$

is a feasible linear programming problem. However, employing the simplex method with slack variables introduced to convert the constraint in equalities into equations, we could not eliminate all of the slack variables from the solution of the programming problem. Hence, we had to resort to a brute force approach of examining all solutions of combinations of three equations selected from the six equations in (16). For each such solution (x, y, r), we defined the residuals

(45) $$\eta_i = |x_i\, x + y_i\, y + u_i\, r - v_i|, \qquad i = 1, \cdots, 6.$$

Then we ordered the residuals

(46) $$w_j = \eta_{i_j}$$

so that

$$w_6 \le w_5 \le w_4 \le w_3 \le w_2 \le w_1.$$

Note that $w_4 \approx w_5 \approx w_6 \approx 0$ to machine accuracy via solution of three of the equations. We considered the resulting three minimization problems. The

(47) $$\min w_1$$

yields the solution (x, y, r) to the linear programming problem (43)-(44),

(48) $$\min w_2$$

yields the solution (x, y, r) corresponding to the best fit of five of the six equations in (16)

and

(49) $$\min w_3$$

yields the solution (x, y, r) corresponding to the best fit of four of the six equations in (16). As noted above fake solutions were observed for the solutions corresponding to the minimization problems (47), (48) and (49).

C. Nonlinear Least Squares for $\bar{u}$. Considering the least squares problem (41), we considered $x = x(\bar{u}), y = y(\bar{u})$ and $r = r(\bar{u})$ as $\bar{u} = (u_1, u_2, u_3, u_4, u_5, u_6)$ is the basic input for the calculation of (x, y, r). We can state the problem as

(50) $$\zeta = \min_{\bar{u}} F(\bar{u}),$$

where

(51) $$F(\bar{u}) = S(x(\bar{u}), y(\bar{u}), r(\bar{u}))$$

and $(x(\bar{u}), y(\bar{u}), r(\bar{u}))$ is the solution of (41)-(42) noting that $v_i = v_i(u_i)$. In effect this can be viewed as modifying the measured data $\bar{u}$ to a nearby consistent data set. The measured data set $\bar{u}$ was taken as the input to a Newton's method iteration for the solution of the equations

(52) $$\frac{\partial F}{\partial u_i}(\bar{u}) = 0, \quad i = 1, \cdots, 6,$$

when the data is good this method works. Shifting to the question of using the best four at the $u_i, i = 1, \cdots, 6$, (50)-(52) can be modified for four of the u_i and coupled to any reasonable linear least squares or linear programming method which produces a set of u_i, u_j, u_k, u_l associated with the optimal (x, y, r) that can be taken as the initial u's for the Newton method iteration for the solution of the modified nonlinear least squares problem. When the solution to the linear method was reasonable and the Newton iteration converged, the calculation cost for the Newton method did not yield much of a gain in solution accuracy. Sometimes the linear solution was reasonable

and the Newton method failed to converge. Gradient procedures to find a good starting set of u's were tried and discarded as were simple search methods. Occasionally, the Newton method would converge to a fake solution especially when the linear procedure produced one.

D. Nonlinear Least Squares for x, y and r. We recall equation (5) and consider the problem

$$v = \min_{x,y,r} G(x, y, y) \tag{53}$$

where

$$G(x, y, r) = \sum_{i=1}^{6} \left(\sqrt{r^2 + r_i^2 - 2x_i x - 2y_{iy}} - r - u_i \right)^2 \tag{54}$$

and its modifications using four and five of the measured data $u_i, i = 1, 2, \cdots, 6$. As in **C.** above, a Newton's method iteration is required to solve the equations

$$\begin{aligned} \frac{\partial G}{\partial x}(x, y, r) &= 0, \\ \frac{\partial G}{\partial y}(x, y, r) &= 0, \\ \frac{\partial G}{\partial r}(x, y, r) &= 0, \end{aligned} \tag{55}$$

and an initial reasonable (x, y, r) is needed to start the Newton iteration. The starting (x, y, r) can be obtained from any of the linear processes mentioned above. As before good data yielded convergence to a good solution, but the improvement in the solution was not significant enough to justify the computational overhead of the Newton method. An attempt was made to consider the method (53)-(55) as a filter for identifying bad data and the exclusion of fake solutions. However, convergence to fake solutions of the nonlinear least squares was observed. So, divergence of the process cannot be taken as a filter for fake solutions.

E. Statistical Estimations. Lots of solutions can be generated when the data $u_i, i = 1, \cdots, 6$, contain error. For example, three equations from six yields 20 different solutions, four equations from six yields via least squares 15 more different solutions, five from 6 yields 6 more and least squares for all six yields another for a total of 42 different solutions which can be seen to coalesce into the correct (x, y, r) as the error in the data tends to zero. There are many ways to

generate an infinite supply of different solutions when the measurement error is now zero. For example, one could use weighted least squares, etc. Focusing on the x coordinate of the solution we obtain from whatever process a set of estimations of x which can be written in the form $x+\varepsilon_i, i=1,\cdots,n$, where n is the number of approximate solutions that you wish to generate and ε_i is the difference between the approximate x and the true x. For good data the average of the approximate x's is a reasonable estimation of the true x. As the data measurement error increases the accuracy of the average degenerates rapidly and the standard deviation increases quickly. Consideration of plots of the approximate solutions in the presence of various synthetic error magnitudes shows the existence of clusters of approximate solutions with one such cluster near the known solution point. For example, see Figures 2-9. So the question of identifying the best cluster of approximations and using the average of coordinates of the points in that cluster as the solution approximation is another process that arises in a natural way.

F. Cluster Analysis. An early observation in the process of identification of the best cluster was that any process of selecting points without reference back to the original measured data could turn rapidly into a random number generator as the data error increased. Reference back to the measured u_i tended to slow down the process of fake solution generation. We shall outline one attempt that had a good chance of being a reasonable process. Let the solution approximations from some process such as least squares, etc. be denoted by $(x_k, y_k, r_k), k=1,\cdots,n$ and let

$$u_{ik} = \sqrt{r_k^2 + r_i^2 - 2x_i x_k - 2y_i y_k} - r_k, i=1,\cdots,6, \tag{56}$$

where here the i subscript refers to the Site #i coordinates while the k subscripts refer to the coordinates/output of the calculation of the kth approximate solution. Let

$$J_{ik} = |u_{ik} - u_i|, \tag{57}$$

where u_i denotes the measurement data as defined above in (17). Let

$$w_{jk} = J_{i_j k} \tag{58}$$

be an ordering of the J_{ik} such that for each k

$$w_{1k} \geq w_{2k} \geq \cdots \geq w_{6k}. \tag{59}$$

Various data fit measures for each solution (x_k, y_k, r_k) can be calculated. For example,

$$
\begin{aligned}
&M_6(k) = w_{1k}, \quad k = 1, \cdots, n, \\
&M_5(k) = w_{2k}, \quad k = 1, \cdots, n, \\
&M_4(k) = w_{3k}, \quad k = 1, \cdots, n, \\
&L_6^2(k) = \left(\sum_{j=1}^{6} w_{jk}^2 \right)^{1/2}, k = 1, \cdots, n, \\
&L_5^2(k) = \left(\sum_{j=1}^{5} w_{jk}^2 \right)^{1/2}, k = 1, \cdots, n, \\
&L_4^2(k) = \left(\sum_{j=1}^{4} w_{jk}^2 \right)^{1/2}, k = 1, \cdots, n, \\
&L_6^1(k) = \left(\sum_{j=1}^{6} w_{jk} \right), k = 1, \cdots, n, \\
&L_5^1(k) = \left(\sum_{j=1}^{5} w_{jk} \right), k = 1, \cdots, n, \text{ and} \\
&L_4^1(k) = \left(\sum_{j=1}^{4} w_{jk} \right), k = 1, \cdots, n,
\end{aligned}
\tag{60}
$$

represent respectively the l_∞ norm on the Euclidean space of vectors $\vec{u} = (u_i, \cdots, u_6)$, the l_∞ semi-norm of the 5th largest absolute value, the l_∞ semi-norm of the 4th largest absolute value, the l_2 norm, the l_2 semi-norm of the smallest to the 5th largest absolute value, the l_2 semi-norm of the smallest to the 4th largest, the l_1 norm, the l_1 semi-norm of the smallest to the 5th largest absolute value, and the l_1 semi-norm of the smallest to the 4th largest. Each norm/data fit measure was ordered large to small and say the bottom 1/5 (could be 1/4 or whatever) were selected and a tally mark placed next to the solutions contained in that position of the list representing a "good data fit" relative to that measure/norm/semi-norm. The approximate solutions with the highest number of tally marks defined the cluster. The approximate solutions in the cluster could be viewed as reasonable approximations in a broad sense. However, as the data error increased fake solutions were observed.

To filter out fake solutions some additional measurements could be very helpful.

IV. An Angular Filter

Currently a short base line interferometer is under development for location at the central Site #0. The digital sample rate is estimated at a voltage measurement per 2 nanoseconds versus a 50 nanoseconds rate for the current LDAR system. The outcome of these measurements would allow the estimation of the direction cosines of the (x, y, z) event location point/vector; i.e., an estimate of azimuth angle θ and angle of elevation ϕ would yield an estimate of

$$\left(\frac{x}{r}, \frac{y}{r}, \frac{z}{r}\right) = (\cos\phi\cos\theta, \cos\phi\sin\theta, \sin\phi) \tag{61}$$

Assuming that these angles could be determined to within $\pm 2^\circ$, tests were run on synthetic data which added error into the true angles selected from a uniform distribution of $\pm 2^\circ$. The solid angle

$$\Omega = \left\{(\theta, \phi): \phi_m - 4^\circ \le \theta \le \theta_m + 4^\circ, \phi_m - 4^\circ \le \phi \le \phi_m + 4^\circ\right\} \tag{62}$$

was designated as the angular filter, where θ_m and ϕ_m denote, respectively, the measured azimuth angle, and the measured elevation angle. Solutions outside of Ω were discarded. Figures 2-5 show the effect of the Ω filter on the 42 least squares solutions. These figures/results show the promise of the filter while its use in practice may differ from that of eliminating approximate solutions to find a good cluster. For example, one may wish to use it to test a solution obtained from some process - say the current method in operational use. Figures 6 and 7 show a good solution cluster excluded by the Ω filter. Use of the filter may incur the cost of exclusion of some good solutions. Figures 8 and 9 show a fake solution cluster, from which the current operational method would generate a solution. Clearly Ω would designate such a solution as a fake. Also, the volume $V(\Omega, r)$ of the tetrahedron with "vertex" at $(0, 0, 0)$ and "height" r which is generated by Ω has a rate of increase $\frac{dV}{dr} \sim r^2$. Thus, for large r, the Ω filter would probably accept some fakes.

V. Conclusions and Open Questions

Conclusions:

A. Any method will work with good measurement data.

B. No fool-proof method for eliminating fake solutions was found.

C. The Ω Filter shows promise of eliminating fakes with the possible cost of excluding some acceptable solutions. This cost might be acceptable given the rate of generation of lightning events within a thunder storm.

Open Questions:

A. Which method is best? "Best" would need careful definition. No attempt was made to find a "Best" method.

B. What is the best way to utilize the Ω filter to minimize acceptance of fakes.

C. What are the possible ways that fakes can be introduced via errors in measurements? Electrical Interference, etc.?

D. The condition number of the various matrices corresponding to the Sites $i, j,$ and k, $1 \le i < j < k \le 6$, needs to be employed in the solution process to reduce error amplification.

E. Design of other mathematical filters or measurement filters to eliminate fakes.

References

1. Cianos, N., Oetzel, G.N., and Pierce, E.T., A technique for accurately locating lightning at close ranges, *J. Appl. Meteor.*, **11**, 1120-1127, 1972.

2. Holle, R.L. and Lopez, R.E., *Overview of real-time lightning detection systems and their meteorological uses,* NOAA Technical Memorandum ERL NSSL-102, 1993.

3. Koshak, W.J., Christian, H.J., and Solakiewicz, R., *On the retrieval of lightning radio sources from time of arrival data,* to appear.

4. MacClement, W.D. and Murty, R.C., VHF direction finder studies of lightning, *J. Appl. Meteor.*, **17**, 786-795, 1978.

5. Maier, L.M. , Lennon, C.L., Britt, T. and Schaefer, S., Lightning Detection and Ranging LLDAR) System Performance Analysis, Proceedings for the Sixth Conference on Aviation Weather Systems at the 75th American Meteorological Society Meeting, Dallas, Texas, January 15-20, 1995.

6. Murty, R.C. and MacClement, W.D., VHF direction finder for lightning location, *J. Appl. Meteor.*, **12**, 1401-1405, 1973.

7. Oetzel, G.N. and Pierce, E.T., VHF technique for locating lightning, *Radio SCI.*, **4**, 199-201, 1969.

8. Poehler, H.A., *An Accuracy Analysis of the LDAR System,* NASA Contractor Report CR-154631, 1977.

9. Poehler, H.A. and Lennon, C.L., *Lightning Detection & Ranging System (LDAR) system description & performance objectives,* NASA Technical Memorandum TM-74105, 1979.

10. Proctor, D.E., A hyperbolic system for obtaining VHF radio pictures of lightning, *J. Geophys. Res.*, **76**, 1478-1489, 1971.

11. Proctor, D.E., VHF radio pictures of cloud flashes, *J. Geophys. Res.*, **86**, 4041-4071, 1981.

12. Proctor, D.E., Lightning and precipitation in a small multicellular thunderstorm, *J. Geophys. Res.*, **88**, 5421-5440, 1983.

13. Proctor, D.E., Regions where lightning flashes began, *J. Geophys. Res.*, **96**, 5099-5112, 1991.

14. Rustan, P.L., Uman, M.A., Childers, D.G., Beasley, W.H., and Lennon, C.L., Lightning source locations from VHF radiation data for a flash at Kennedy Space Center, *L. Geophys. Res.*, **85**, 4893-4903, 1980.

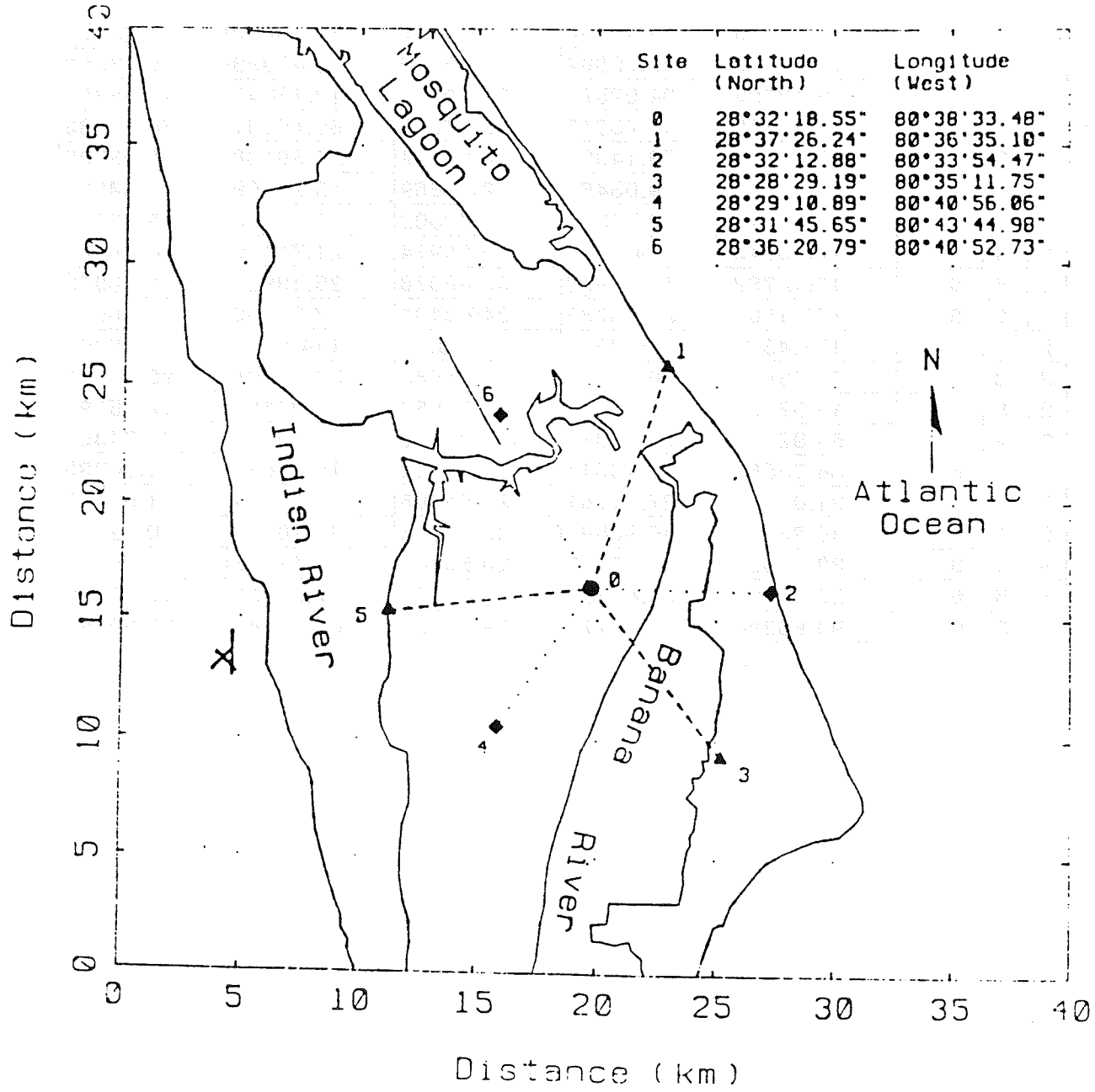
15. Taylor, W.L., A VHF technique for space-time mapping of lightning discharge processes. *J. Geophys. Res.*, **83**, 3575-3583, 1978.

16. Thomson, E.M., Medelius, P.J., and Davis, S., A system for locating the sources of wideband dE/dt from lightning, *J. Geophys. Res.*, **99**, 22, 793-22,802, 1994.

17. Twomey, S.A., *Introduction to the Mathematics of Inversion in Remote Sensing and Indirect Measurements,* Elsevier, New York, 1977.

18. Young, D.M., and Gregory, R.T., *A survey of Numerical Mathematics, Volumes I & II,* Addison Wesley Publishing Company, Reading Massachusetts, 1972.

FIGURE 1: LOCATIONS OF LDAR SITES.



TABLE#1:	L-00 CONDITION NUMBER FOR THE MATRIX OF (19)				
	WITH THETA= 45 DEGREES AND R=50 KM				
SITES: I, J, K	PHI=0 DEG'S	PHI=18 DEG'S	PHI=36 DEG'S	PHI=54 DEG'S	PHI=72 DEG'S
1 , 2 , 3	61.0435	55.93494	45.16827	34.93892	27.20576
1 , 2 , 4	146.0429	99.75604	52.32759	30.47398	20.38362
1 , 2 , 5	39.43026	34.02578	24.29061	16.73847	11.98991
1 , 2 , 6	52.37815	54.23213	59.37658	65.49181	66.27332
1 , 3 , 4	100.7596	76.14307	44.83299	27.66138	18.98708
1 , 3 , 5	32.64875	29.03469	21.9869	15.95798	11.84311
1 , 3 , 6	21.07241	20.74303	19.70602	17.8936	15.38053
1 , 4 , 5	178.6548	94.3098	40.25974	21.71146	14.14445
1 , 4 , 6	1203.788	177.7992	52.06078	25.43921	16.0372
1 , 5 , 6	152.1562	244.3768	349.3181	77.37132	41.66919
2 , 3 , 4	131.4395	174.2611	1930.4	144.6241	75.6265
2 , 3 , 5	70.76457	60.91602	43.5988	30.80794	25.34381
2 , 3 , 6	13.38127	13.341	13.22082	13.02164	12.75794
2 , 4 , 5	59.92292	53.0831	39.99109	29.05757	21.74881
2 , 4 , 6	34.35554	31.10017	24.43622	18.38688	14.06935
2 , 5 , 6	31.03037	27.61281	20.98866	15.37501	11.58672
3 , 4 , 5	35.88189	35.36891	33.79395	31.20216	30.69836
3 , 4 , 6	23.17594	22.47862	20.60585	18.11648	17.20661
3 , 5 , 6	22.03605	21.42122	19.75133	17.59359	16.77516
4 , 5 , 6	99.83394	93.51775	78.98998	63.55145	50.92098

TABLE#2:	L-00 CONDITION NUMBER FOR THE MATRICES OF (19)				
	WITH THETA=45 DEGREES AND PHI=36 DEGREES				
SITES: I, J, K	R=10 KM	R=20 KM	R=30 KM	R=40 KM	R=50 KM
1 , 2 , 3	7.562704	16.79023	26.36976	35.796	45.16827
1 , 2 , 4	6.207406	17.32838	29.06532	40.72616	52.32759
1 , 2 , 5	3.800319	8.940697	14.12844	19.229	24.29061
1 , 2 , 6	13.99799	24.45105	35.80679	47.52315	59.37658
1 , 3 , 4	5.968515	15.2238	25.18018	35.03581	44.83299
1 , 3 , 5	3.606738	8.129786	12.81055	17.41429	21.9869
1 , 3 , 6	3.98372	7.696629	11.68949	15.69356	19.70602
1 , 4 , 5	5.305135	12.67618	21.85456	31.064	40.25974
1 , 4 , 6	7.163752	15.96919	28.01849	40.06938	52.06078
1 , 5 , 6	17.71435	80.49199	169.3603	260.5127	349.3181
2 , 3 , 4	37.80313	230.6574	1426.055	4403.898	1930.4
2 , 3 , 5	8.939254	17.56767	26.33672	34.99567	43.5988
2 , 3 , 6	3.54418	5.862628	8.291217	10.74992	13.22082
2 , 4 , 5	8.244946	16.25348	24.13949	32.07376	39.99109
2 , 4 , 6	4.713811	9.528722	14.48196	19.45647	24.43622
2 , 5 , 6	4.326795	8.541748	12.7215	16.86451	20.98866
3 , 4 , 5	9.347797	15.01976	21.15089	27.45198	33.79395
3 , 4 , 6	4.923535	8.496019	12.49855	16.54372	20.60585
3 , 5 , 6	4.831757	8.3426	12.12114	15.93046	19.75133
4 , 5 , 6	14.61789	29.38895	45.39818	62.07432	78.98998

TABLE#3:	L-00 CONDITION NUMBER LEAST SQUARES MATRICES FROM (19)				
	WITH THETA=45 DEGREES AND R=50 KM				
SITES: I, J, K	PHI=0 DEG'S	PHI=18 DEG'S	PHI=36 DEG'S	PHI=54 DEG'S	PHI=72 DEG'S
1 , 2 , 3	2987.496	2388.022	1569.014	950.2631	585.7231
1 , 2 , 4	15506.6	6879.296	1673.905	580.0257	265.5546
1 , 2 , 5	3733.733	2650.927	1215.24	599.3609	321.1228
1 , 2 , 6	1930.539	1949.161	2343.786	2862.446	2940.187
1 , 3 , 4	5998.944	3442.668	1209.967	469.6478	225.836
1 , 3 , 5	953.9588	719.7244	354.0826	189.247	107.3342
1 , 3 , 6	629.783	612.8223	560.3766	473.0363	362.5532
1 , 4 , 5	104203	27931.26	4473.017	1286.358	577.4056
1 , 4 , 6	2649161	58201.13	5091.993	1253.955	517.4218
1 , 5 , 6	18575.84	45718.49	84247.99	4246.766	1271.488
2 , 3 , 4	12363.5	20597.47	2495624	13788.87	2983.77
2 , 3 , 5	4428.875	3280.984	1681.046	826.7447	452.4155
2 , 3 , 6	377.0717	366.1686	335.1751	289.021	234.8757
2 , 4 , 5	3358.238	2625.204	1473.79	765.6875	421.2037
2 , 4 , 6	1182.481	916.8083	474.3499	265.4693	153.7965
2 , 5 , 6	2123.316	1679.989	969.1752	520.407	297.7738
3 , 4 , 5	1672.863	1557.485	1280.792	1092.266	849.8152
3 , 4 , 6	433.0106	409.7616	350.7537	278.268	210.0583
3 , 5 , 6	343.2147	327.5543	287.0623	235.659	185.2436
4 , 5 , 6	11024.89	9260.442	5741.057	3450.198	2245.19

TABLE#4:	L-00 CONDITION NUMBER FOR LEAST SQUARES MATRICES FROM (19)				
	WITH THETA=45 DEGREES AND PHI=36 DEGREES				
SITES: I, J, K	R=10 KM	R=20 KM	R=30 KM	R=40 KM	R=50 KM
1 , 2 , 3	37.05391	215.527	533.1044	984.2259	1569.014
1 , 2 , 4	26.6828	190.8348	523.6204	1018.672	1673.905
1 , 2 , 5	38.41716	182.0089	428.0012	772.625	1215.24
1 , 2 , 6	130.5545	395.334	849.6547	1499.456	2343.786
1 , 3 , 4	19.31481	139.8363	381.1006	738.2283	1209.967
1 , 3 , 5	10.38426	51.70143	122.7587	223.234	354.0826
1 , 3 , 6	18.69834	83.11837	194.7259	353.7285	560.3766
1 , 4 , 5	72.1908	500.4724	1389.842	2716.575	4473.017
1 , 4 , 6	68.01589	541.2159	1553.214	3074.353	5091.993
1 , 5 , 6	251.0914	4836.73	20463.48	47424.2	84247.99
2 , 3 , 4	694.2905	31356.67	1272577	1.27E+07	2495624
2 , 3 , 5	51.54088	246.6884	586.2598	1064.749	1681.046
2 , 3 , 6	12.89667	51.93632	118.5101	212.9032	335.1751
2 , 4 , 5	51.72648	229.1964	527.2562	942.3237	1473.79
2 , 4 , 6	15.03352	67.72181	160.5597	296.5604	474.3499
2 , 5 , 6	35.81037	152.6888	348.2725	620.553	969.1752
3 , 4 , 5	100.3142	259.6986	503.1655	840.332	1280.792
3 , 4 , 6	16.43427	57.55901	126.9724	224.7061	350.7537
3 , 5 , 6	15.0548	49.47289	106.3153	185.5214	287.0623
4 , 5 , 6	219.3776	835.2269	1955.032	3589.375	5741.057

FIGURE 2: ADMISSIBLE SOLUTIONS FROM 42 LEAST SQUARES SOLUTIONS WITH MAXIMUM RANDOM ERROR 750 METERS AND SOLUTION AT X=15 KM, Y= 15 KM , AND Z=20 KM

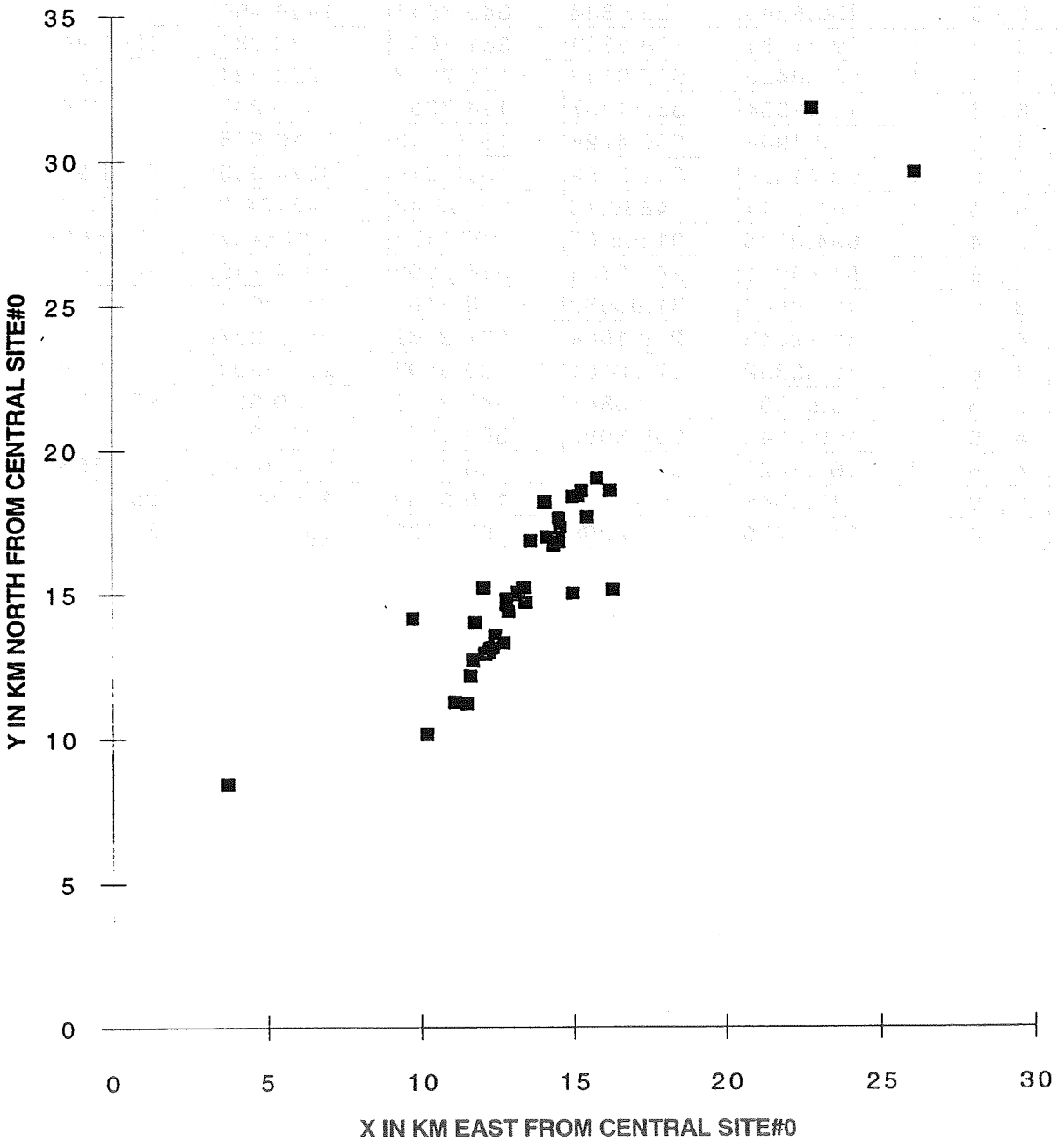


FIGURE 3: ANGLE FILTERED SOLUTIONS FROM ADMISSIBLE SOLUTIONS WITH MAXIMUM RANDOM ERROR 750 METERS AND SOLUTION AT X=15 KM, Y= 15 KM AND Z= 20 KM

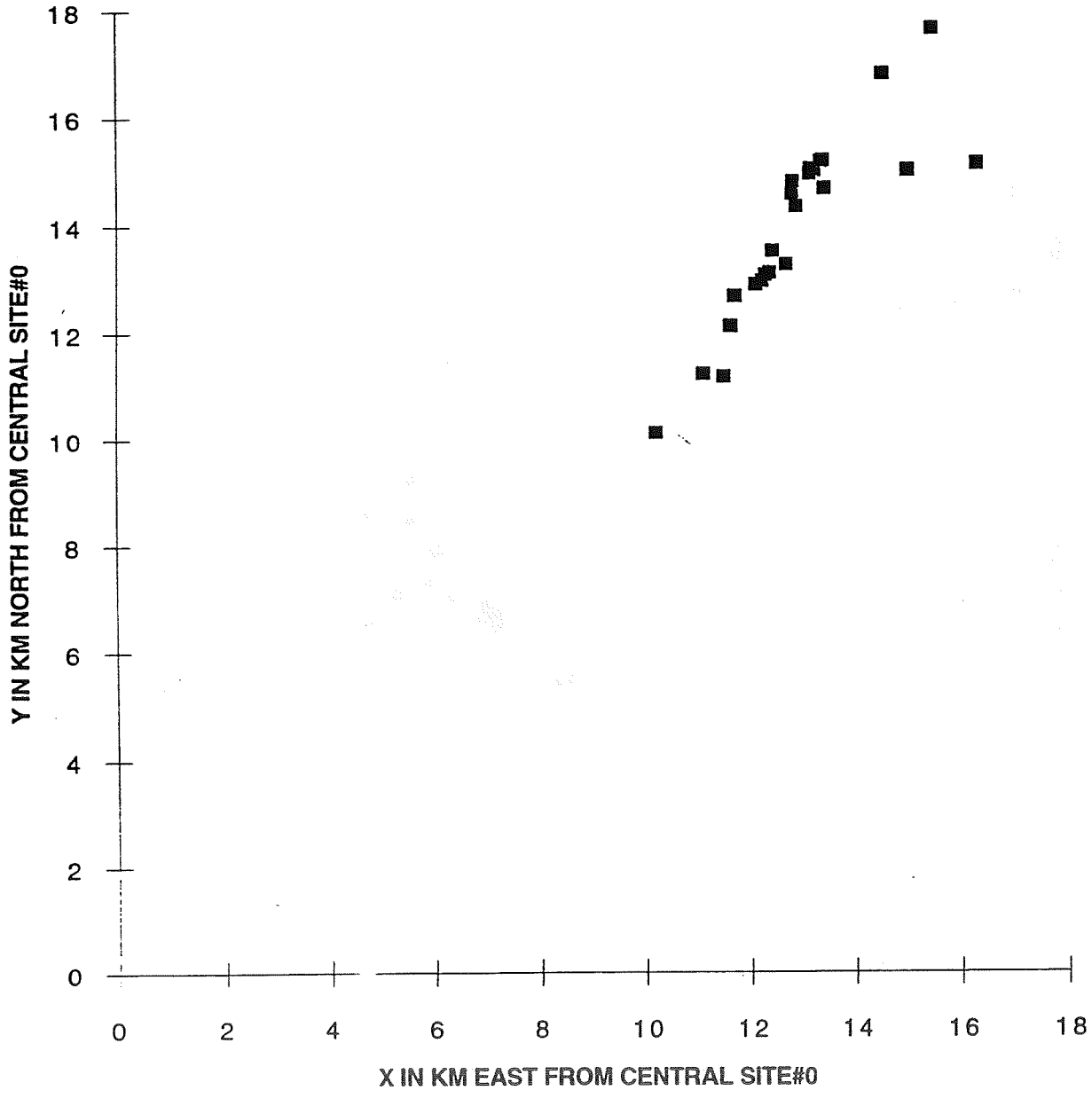
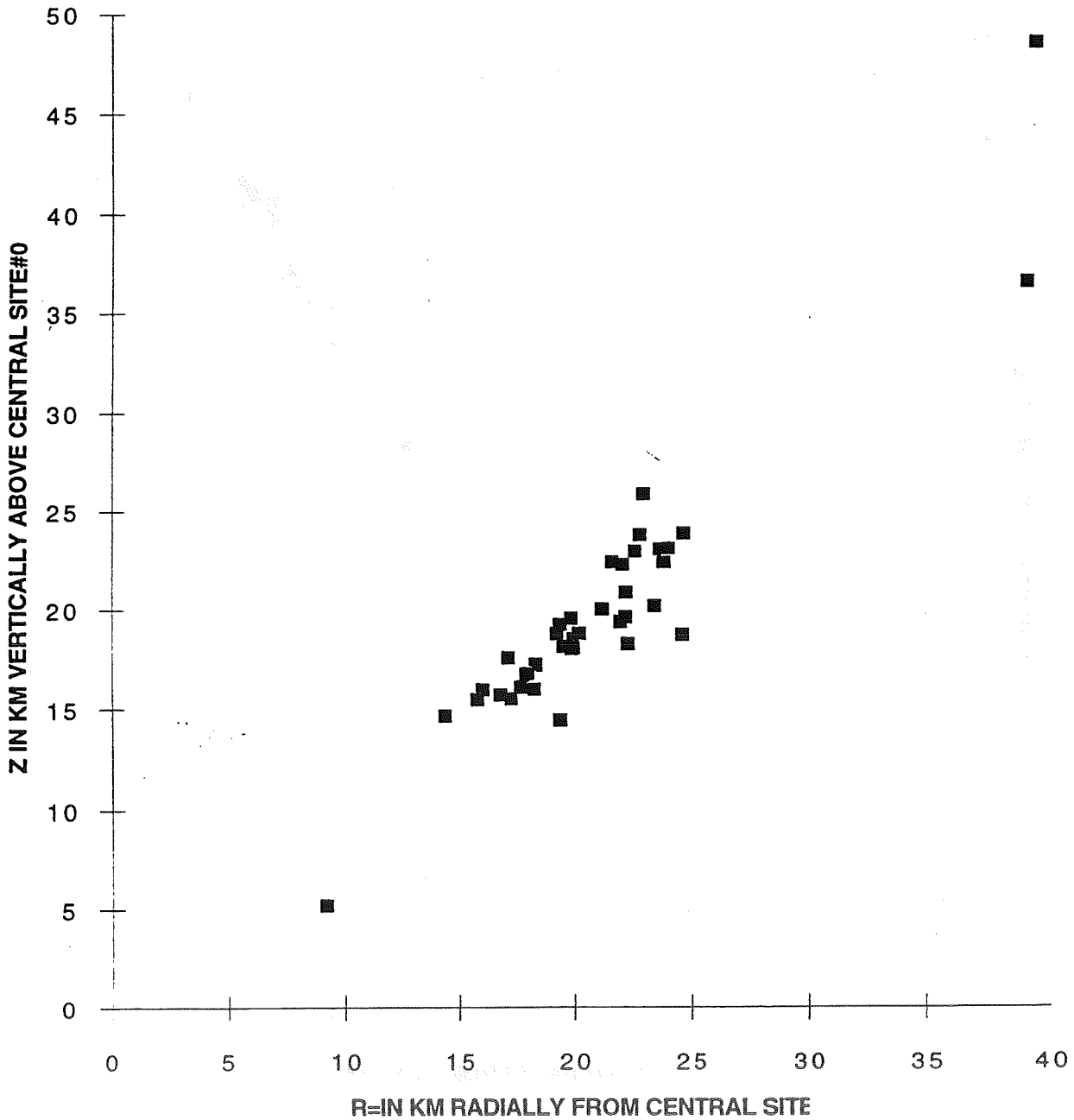


FIGURE 4: ADMISSIBLE SOLUTIONS FROM 42 LEAST SQUARES SOLUTIONS WITH MAXIMUM RANDOM ERROR 750 METERS AND SOLUTION AT X=15 KM, Y=15 KM, AND Z= 20 KM



R=SQR(X^2+Y^2)

FIGURE 5: ANGLE FILTERED SOLUTIONS FROM ADMISSIBLE SOLUTIONS WITH MAXIMUM RANDOM ERROR 750 METERS AND SOLUTION AT X=15 KM, Y=15 KM, AND Z= 20 KM

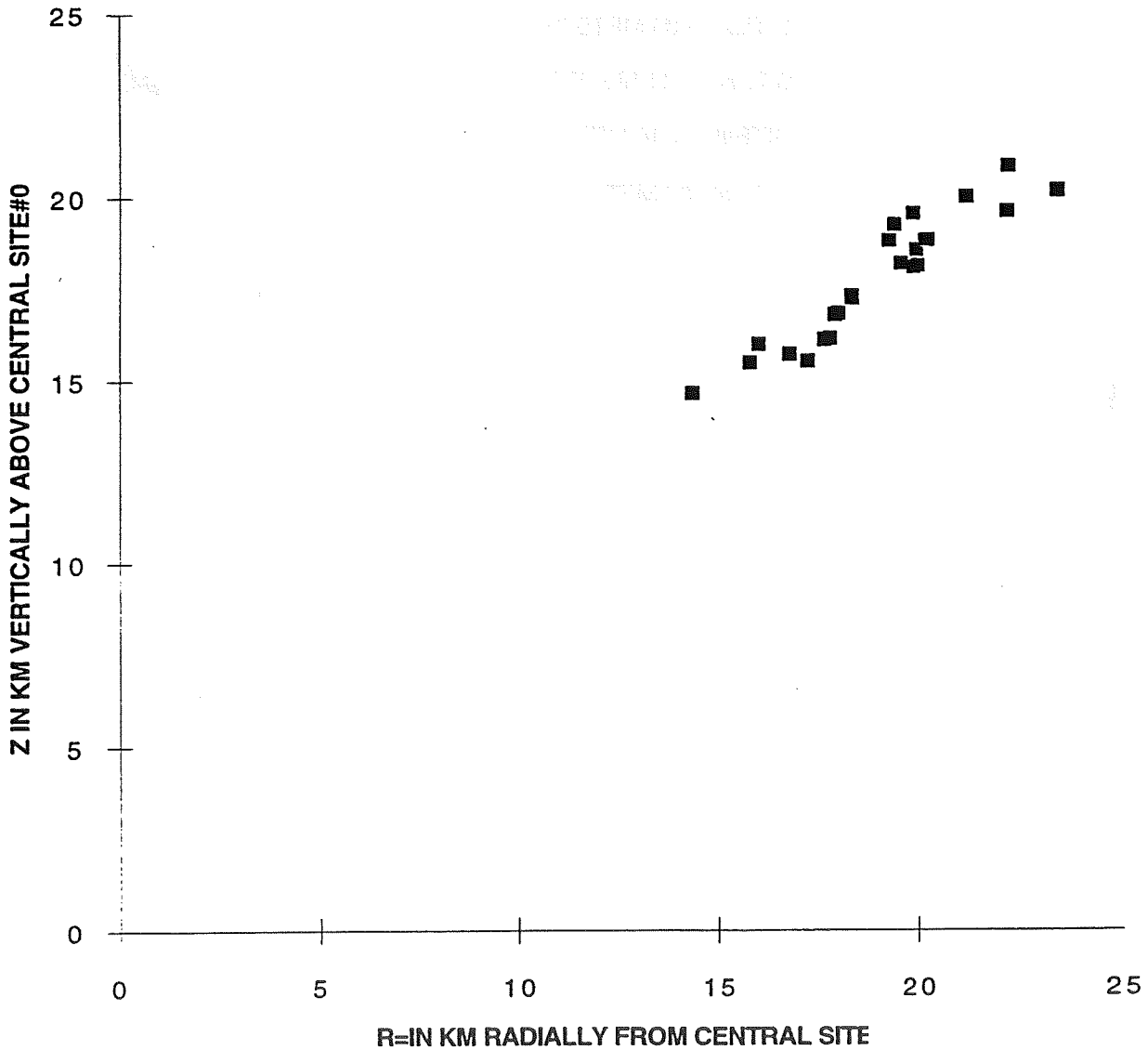


FIGURE 6: ADMISSIBLE SOLUTIONS FROM 42 LEAST SQUARES SOLUTIONS WITH SOLUTION AT X= 15 KM, Y=15 KM, AND Z=20 KM

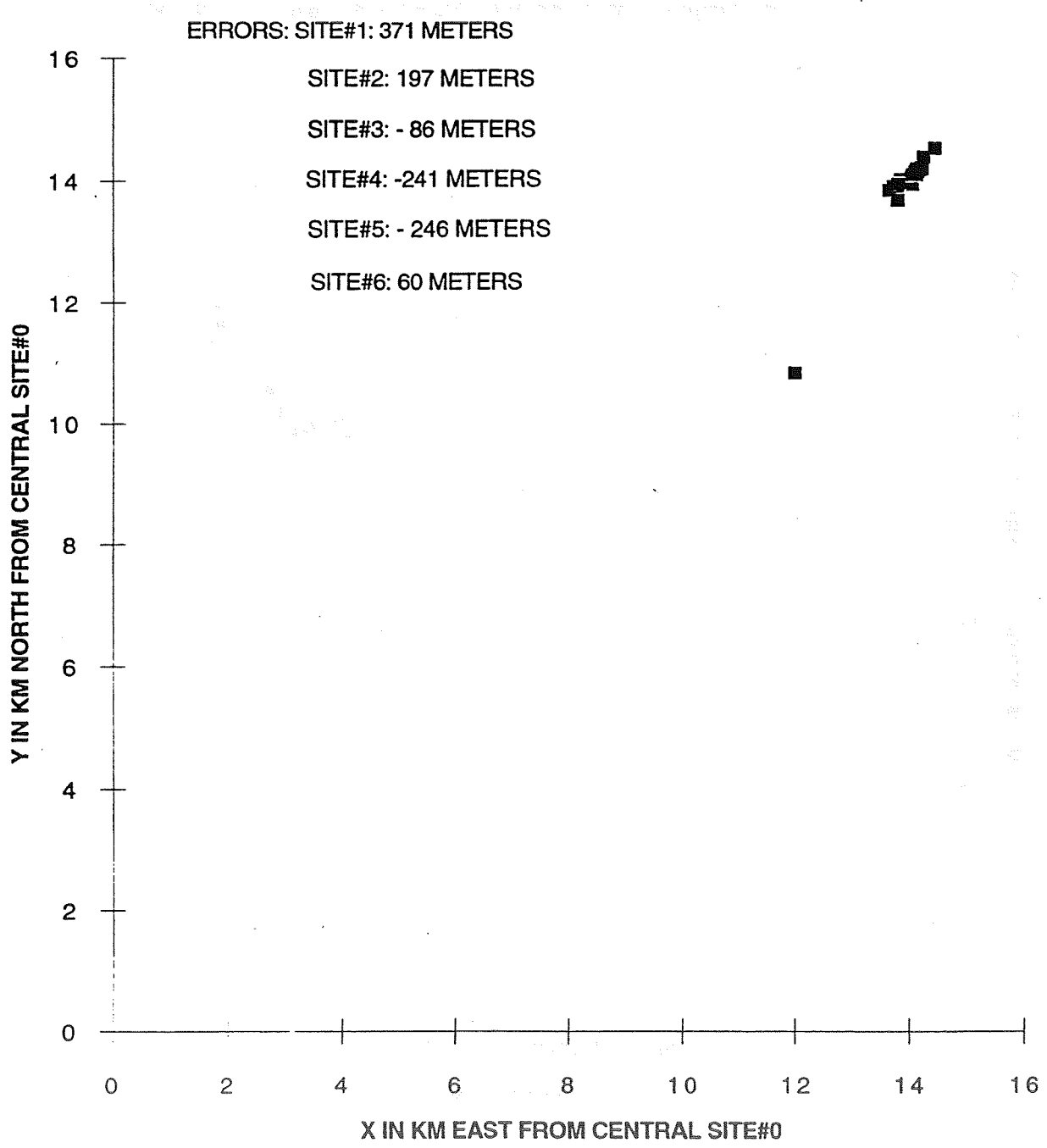


FIGURE 7: ADMISSIBLE SOLUTIONS FROM 42 LEAST SQUARES SOLUTIONS WITH SOLUTION AT X=15 KM, Y=15 KM, AND Z=20 KM

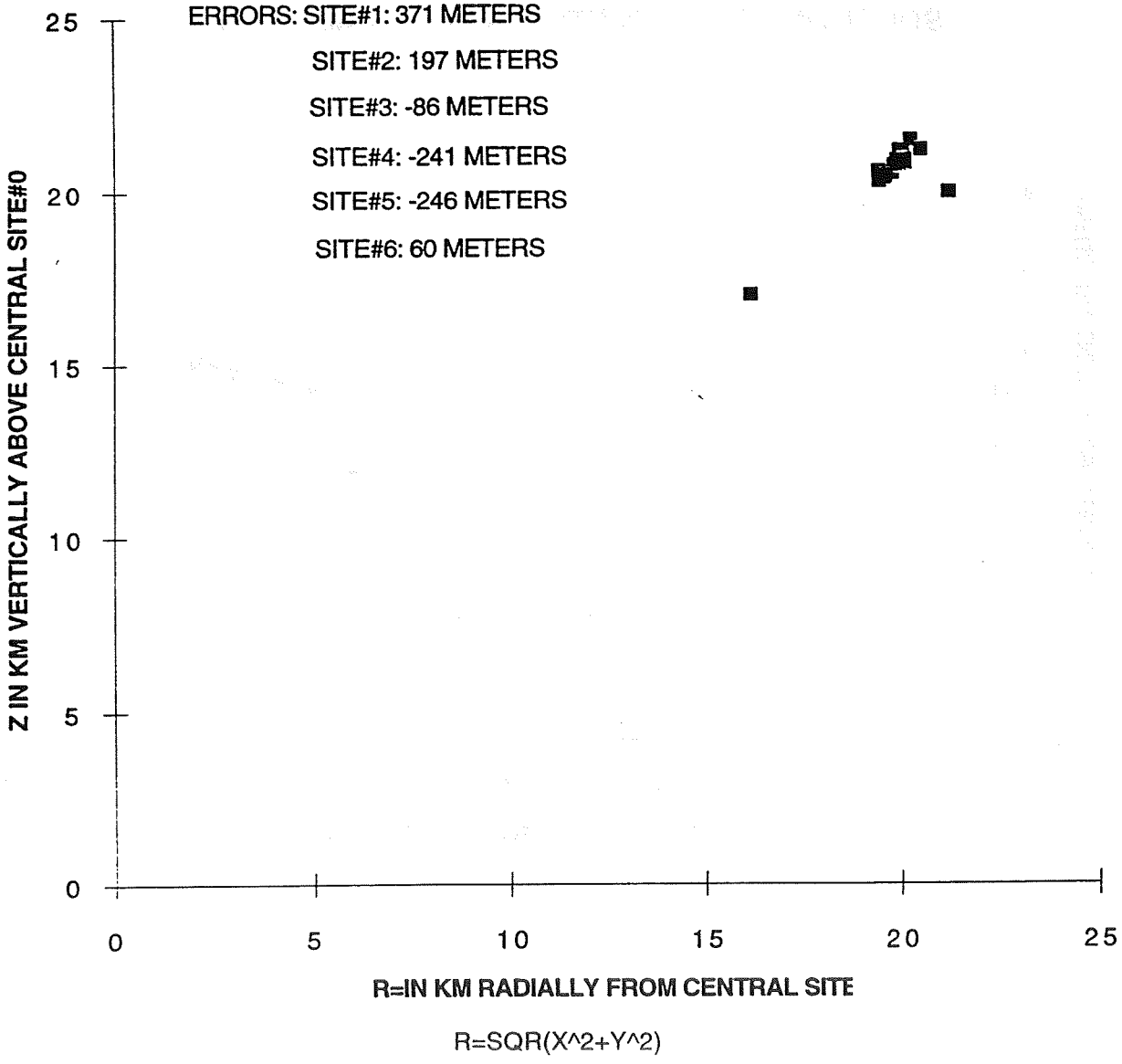


FIGURE 8: ADMISSIBLE SOLUTIONS FROM 42 LEAST SQUARES SOLUTIONS WITH SOLUTION AT X=15 KM, Y=15 KM, Z=20 KM

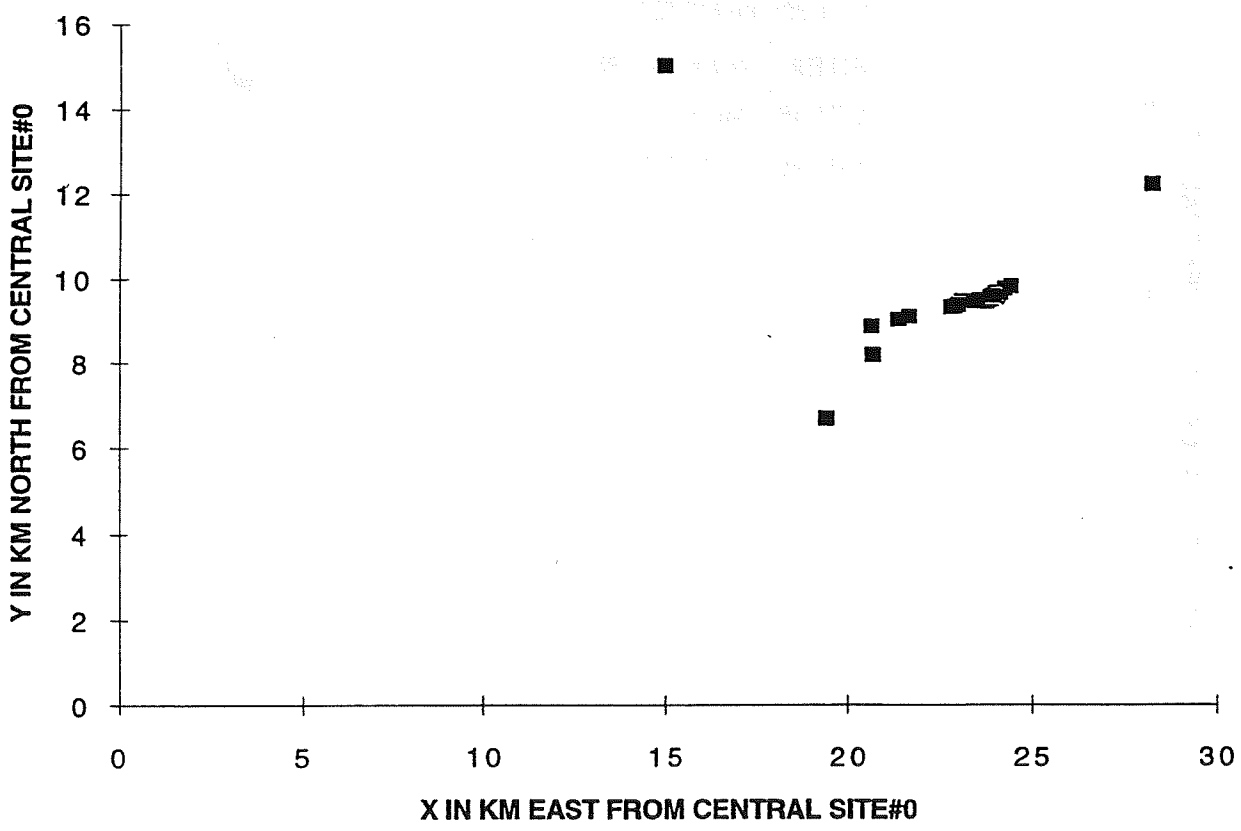
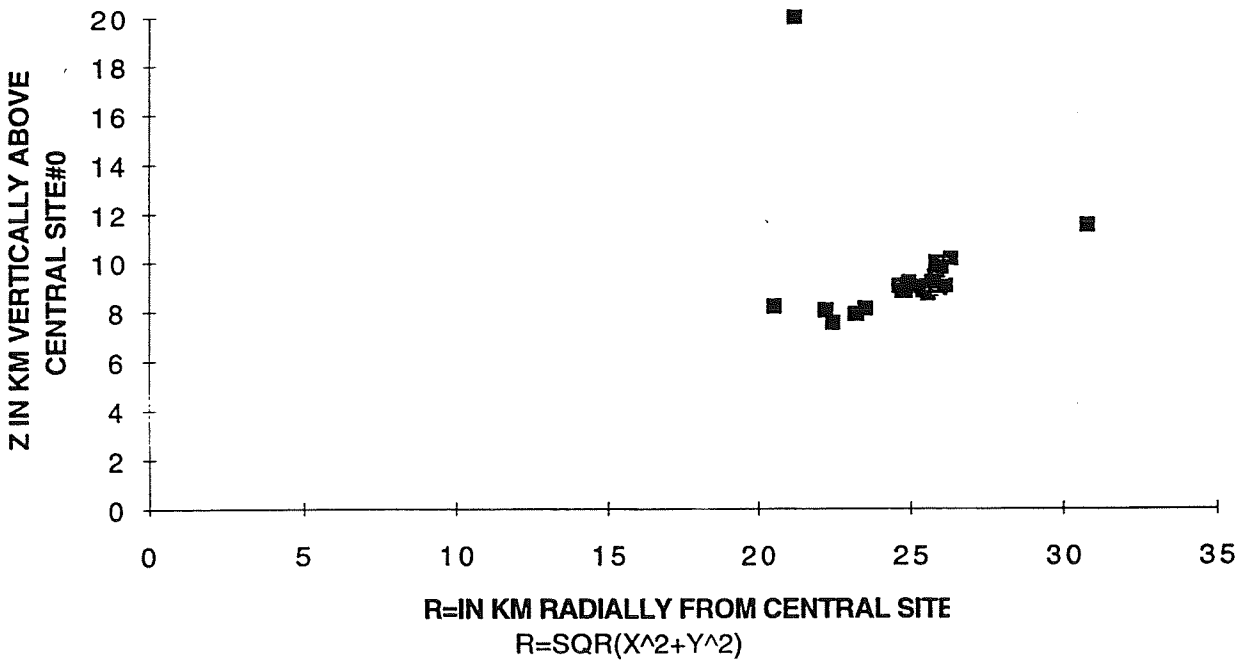


FIGURE 9: ADMISSIBLE SOLUTIONS FROM 42 LEAST SQUARES SOLUTIONS WITH SOLUTION AT X=15 KM, Y=15 KM, AND Z=20 KM



1995 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

54-51
7744
P-34

ON-LINE REMOVAL OF VOLATILE FATTY ACIDS FROM CELSS ANAEROBIC BIOREACTOR VIA NANOFILTRATION

Dr. Guillermo Colon
Associate Professor
Chemical Engineering Department
University of Puerto Rico
Mayaguez, Puerto Rico

KSC Colleague - John C. Sager
Life Sciences

Contract Number NASA-NGT-60002
Supplement 19

August 9, 1995

PAGE 86 INTENTIONALLY BLANK

ACKNOWLEDGMENTS

I would like to express my appreciation for being selected to participate in the 1995 NASA/ASEE Summer Fellowship Program. To Dr. E. Ramon Hosler (University of Central Florida) and to Mr. Gregg Burkingham (NASA University Manager) my congratulations, for their capable administration of the program.

I also want to thank my NASA colleague Dr. John C. Sager for his invaluable support and friendship during this summer. Special thanks are extended to Mr. Barry Finger, Mr. Michael Alazraki and Dr. Richard F. Strayer for their continuous support and arranging of the laboratory facilities to carry out the experimental work reported in this document. Thanks to Barbara Peterson and Teresa Englert for their support with the analyses of VFA, inorganic and TOC and to Mr. Jeff Ruddell for his continuous computer assistance.

My gratitude goes also to Ms. Kari Stiles for her dedication and invaluable assistance during my stay in KSC, and to Dr. Raymond Wheeler for his friendship and numerous rides from Titusville to NASA -KSC facilities.

In general I made lots of friends in NASA Kennedy Space Center that I will never forget.

ABSTRACT

The CELSS resource recovery system, which is a waste processing system, uses aerobic and anaerobic bioreactors to recover plants nutrients and secondary foods from the inedible biomass. The anaerobic degradation of the inedible biomass by means of culture of rumen bacteria, generates organic compounds such as volatile fatty acids (acetic, propionic, butyric, VFA) and ammonia. The presence of VFA in the bioreactor medium at fairly low concentrations decreases the microbial population's metabolic reactions due to end-product inhibition. Technologies to remove VFA continuously from the bioreactor are of high interest.

Several candidate technologies were analyzed, such as organic solvent liquid-liquid extraction, adsorption and/or ion exchange, dialysis, electrodialysis, and pressure driven membrane separation processes. The proposed technique for the on-line removal of VFA from the anaerobic bioreactor was a nanofiltration membrane recycle bioreactor. In order to establish the nanofiltration process performance variables before coupling it to the bioreactor, a series of experiments were carried out using a 10,000 MWCO tubular ceramic membrane module. The variables studied were the bioreactor slurry permeation characteristics, such as, the permeate flux, VFA and nutrient removal rates as a function of applied transmembrane pressure, fluid recirculation velocity, suspended matter concentration, and process operating time.

Results indicated that the permeate flux, VFA and nutrients removal rates are directly proportional to the fluid recirculation velocity in the range between 0.6 to 1.0 m/s, applied pressure when these are lower than 1.5 bar, and inversely proportional to the total suspended solids concentration in the range between 23,466 to 34,880. At applied pressure higher than 1.5 bar the flux is not more linearly dependent due to concentration polarization and fouling effects over the membrane surface. It was also found that the permeate flux declines rapidly during the first 5 to 8 hours, and then levels off with a diminishing rate of flux decay.

SUMMARY

NASA has been studying the viability of a controlled ecological life support system (CELSS) for long term space missions for several years. The purpose of this system is to provide all the basic human needs required for life support (food, water, oxygen). Previous studies have been carried out using aerobic and anaerobic bioreactors to recover plant nutrients and secondary foods from the inedible biomass. The anaerobic degradation of the inedible biomass by mean of culture of rumen bacteria , generates organic compounds such as volatile fatty acids (VFA) and ammonia. The presence of VFA in the bioreactor medium at fairly low concentration decreases the microbial metabolic reactions due to end-product inhibition. Technologies to remove VFA from the anaerobic bioreactor is therefore of high interest.

The technique proposed for the on-line removal of VFA from the anaerobic bioreactor broth was a nanofiltration membrane recycle bioreactor. In order to establish the nanofiltration process performance variables before coupling it to the anaerobic bioreactor a series of experiments were carried out using a 10,000 MWCO tubular ceramic membrane module. The variables studied were the bioreactor slurry membrane permeation characteristics such as, the permeate flux, VFA and nutrients removal rates as a function of applied transmembrane pressure, fluid recirculation velocity, suspended matter content, and process operating time.

Results show that the permeate flux, VFA and nutrients removal rates are directly proportional to the fluid recirculation velocity in the range between 0.6 to 1.0 m/s, applied pressure when these are lower than 1.5 bar, and inversely proportional to the bioreactor effluent total suspended solids concentration. The operation at applied pressure higher the 1.5 bar, the permeate flux was not more linearly dependent on applied pressure. The asymptotic flux-pressure relationship observed at high applied pressure is due to concentration polarization and fouling effects over the membrane surface. The permeate flux decreased exponentially with the total suspended solids concentration due to the increase of medium viscosity and polarization and gel layers hydraulic resistances. The VFA and nutrients removal rates were found to be proportional to the total fluid filtration rate and component composition in the bioreactor effluent. It was also found that the permeate flux declined rapidly during the first 5 to 8 hours and then leveled off with a diminishing rate of decay. This phenomena occurred as consequence of the accumulation of suspended matter over the membrane surface which no longer participates in the mass-transport to or away from the membrane.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

ABBREVIATIONS, ACRONYMS AND NOMENCLATURE

BPC = Biomass Production Chamber
CELSS = Controlled Ecological Life Support System
CSTR = Continuous Stirred Tank Reactor
g = gram
hr = hour
L = liter
LMWS = low molecular weight solute
MRB = membrane recycle bioreactor
MW = molecular weight
MWCO = molecular weight cut-off
NF = nanofiltration
TMP = transmembrane pressure
TSS = total suspended solids
VSS = volatile suspended solids
VFA = volatile fatty acid
H_2SO_4 = sulphuric acid
NH_4 = ammonium
PO_4 = phosphate
Fe = iron
Ca = calcium

A = experimental constant
A_w = water membrane permeability constant, L/hr-m^2
C_B = bulk solute concentration, mg/L or mMole
C_p = permeate solute concentration, mg/L or mMole
C_w = membrane wall solute concentration, mg/L or mMole
D = molecular diffusivity, m^2/s
d_H = hydraulic diameter, m
k = polarization boundary layer mass transfer coefficient, m/s
J_V = permeate flux, L/hr-m^2
R_{ads} = adsorbed solute hydraulic resistance, bar-s^2/kg
R_{cp} = concentration polarization hydraulic resistance, bar-s^2/kg
R_g = gel layer hydraulic resistance, bar-s^2/kg
R_m = intrinsic membrane hydraulic resistance, bar-s^2/kg
U, u = fluid recirculation velocity, m/s
α = experimental constant
β = experimental constant
ΔP_m = transmembrane pressure, bar
$\Delta\pi$ = transmembrane osmotic pressure, bar
μ = viscosity, kg/m-s
ρ = density, kg/m^3

I. INTRODUCTION

The National Aeronautical and Space Administration (NASA) has been studying the viability of Controlled Ecological Life Support System (CELSS) for long term space missions for several years. The purpose of this system is to provide all the basic human needs required for life support. The CELSS breadboard project is a large-scale integration system with two main components, a full-scale Biomass Production Chamber (BPC) and the Resource Recovery system. The BPC is a closed system where plants are grown under controlled hydroponic conditions. The plants in CELSS generate oxygen, purified water, and produce edible foods. The resource recovery system which is a waste processing unit uses aerobic and anaerobic bioreactor processes to recover plants nutrients and secondary foods from the inedible biomass. The anaerobic bioreactor is actually operated in a feed and bleed mode of operation. The bioreactor is fed with a microbial population obtained from the cow's rumen together with the inedible portion of plants. Anaerobic degradation of inedible biomass, generates organic compounds such as volatile fatty acids (acetic, propionic, butyric, VFA) and ammonia. The presence of VFA in the fermentation medium at fairly low concentrations decrease the microorganism's metabolic reaction rate due to end-product inhibition. Technologies to remove VFA from the anaerobic bioreactor are therefore of high interest.

1.1 BACKGROUND

Anaerobic digestion a is naturally occurring biological process in which organic material is converted to reduced end products. The anaerobic decomposition of solid organic wastes proceeds in two consecutive steps, liquification and methanogenesis (1, 2). In the first step, microorganisms hydrolyze various components of the solid matter into smaller soluble molecules likes sugars and amino acids. These soluble products are adsorbed by the microbes and are rearranged into CO_2 and volatile fatty acids (acetic, propionic, butyric). The first phase of the anaerobic digestion is known as the acidogenic phase. In the second phase a second group of organisms, the methanogens, complete the digestion process converting the VFA to highly reduced CH_4 and CO_2. Manipulation of bioreactor conditions allow for higher production of VFA and less production of methane. Previous studies have indicated that these acids are phytoxic at certain concentrations (3). VFA at fairly low concentrations in the fermentation medium cause inhibition of the degradation of complex plant polymers, such as cellulose and hemicellulose (4). In order to obtain higher anaerobic conversions of the inedible biomass, it is important to remove VFA to non-toxic levels. One way to remove continuously the VFA from the anaerobic bioreactor fermentation broth is by mean of a membrane recycle bioreactor (MRB).

Basically a membrane recycle bioreactor (MRB) is a reaction vessel, operated as a stirred tank reactor, is coupled in a semi closed-loop configuration via suitable pump to a membrane module containing the appropriate semi-permeable membrane. The membrane should be chosen to retain the microorganisms, enzymes, and most macromolecules while minimizing retention of end-products that affect microorganisms metabolic reactions. For example, in the production of ethanol by yeast, alcohol productivity at 6% (v/v) is only half that at the zero alcohol level, while

it is only 1/100 at 12% (5). Gerhardt and coworkers performed several pioneering experiments on the use of *in vitro* dialysis culture systems for a variety of applications (6). They showed that continuous removal of metabolic products would result in a superior fermentation process. The ethanol productivity using MRB was found to be 30-50 times higher than batch ethanol fermentation (5). The production of propionic acid utilizing culture of Propionibacterium acid-propionici, using renewable resources has been found to be a slow rate fermentation process due to strong product inhibition (7). The use of membrane filter systems for cell recycling improved greatly the fermentation process, allowing to maintain a continuous high-cell-density culture by removing inhibitory metabolic products (8). Propionic acid productivity was 17 times greater than obtained using a continuous stirred tank bioreactor (CSTR). Xavier et al. (9) studied the lactic acid production with cell recycling on an ultrafiltration tubular membrane reactor. They found that the volumetric productivity (36 g/L-hr) was larger than the traditional batch fermentation (3 g/L-hr), or others continuous systems, i.e., CSTR (6 g/L-hr) and immobilized bioreactor (20 g/L-hr).

Nanofiltration (NF) is a rate governed process in which the pressure is the driving force. The feed solution containing macromolecular solutes is introduced into the membrane separator where the solvent and certain microsolutes pass through a semi-permeable membrane and is collected as nanofiltrate. Both theoretically and practically, NF offers an attractive alternative to a number of unit operations in food processing, chemical processing, pharmaceutical, and biotechnological industries. The major limiting step in the use of pressure driven membrane processes (such as NF), particularly with multicomponents feed streams, is the decline of permeate flux as a function time caused by "fouling" of the membrane (10, 11). The decline of flux in NF of a solution or suspension is attributed to concentration polarization and fouling phenomena (such as adsorption, pore blocking, and deposition of solidified solutes, a long term and more or less irreversible process).

During nanofiltration of macromolecular solutes, the linear relationship between permeate flux (J_v) and applied transmembrane pressure (ΔP_m) does not hold very well. At this point solutes get accumulated near the membrane causing permeate flow resistance to rise. The flux become independent of pressure at higher values of ΔP_m. This gives rise to the phenomenon of concentration polarization (10, 12). The steady state mass balance based on the relative rates of solutes transport to the membrane surface by convection and back diffusion, give the following equation for the region where the permeate flux is practically independent on pressure (13):

$$J_v = k \ln\left[\frac{C_w - C_p}{C_B - C_p}\right] \qquad (1)$$

The polarization boundary layer mass transfer coefficient, k, is correlated to the solution physical properties, flow channel dimensions and operating parameters exist in the literature (14, 15). As the solutes bulk concentration increases, the permeate flux decreases exponentially. Using the π

theorem, one can obtain a general correlation of the form:

$$\frac{k\ d_h}{D} = A\left(\frac{\rho\ U\ d_H}{D}\right)^{\alpha}\left(\frac{\mu}{\rho\ D}\right)^{\beta} \qquad (2)$$

As can be observed in the above equation the mass transfer coefficient depends on fluid recirculation velocity (U), solution physical properties (μ, ρ, D), and flow channel dimensions (d_H). The boundary layer mass transfer coefficient is proportional to the fluid velocity, which mean that increasing the fluid velocity reduces greatly the concentration polarization layer.

The water flux through the porous membrane can be describe by Darcy's law:

$$J_v = \frac{\Delta P_m}{R_m \mu_w} = A_w\ \Delta P_m \qquad (3)$$

where R_m, is the intrinsic hydraulic membrane resistance and A_w, is the water membrane permeability constant. It is a function of pore size, tortuosity, membrane thickness, porosity and inversely proportional to fluid viscosity. The permeate flux through the membrane is directly proportional to the applied transmembrane pressure when there is not evidence of concentration polarization and fouling over the membrane surface. From the concentration-polarization model it is evident that the concentration polarization on the membrane polarization will be considerably higher than that in the bulk. This high concentration on one side of the membrane and the very low concentration on the other side create an osmotic pressure difference ($\Delta\pi$) which acts opposite to the applied pressure (ΔP_m). So in Eq. 3, ΔP_m is now replaced by the effective pressure ($\Delta P_m - \Delta\pi$),

$$J_v = \frac{\Delta P_m - \Delta\pi}{R_m \mu_s} \qquad (4)$$

However, this equation is still not adequate to predict the flux. Solutes may get adsorbed, which also cause fouling and also form a distinct polarized layer. These phenomena are accommodated by introducing the resistance in series model (16),

$$J_v = \frac{\Delta P_m - \Delta\pi}{\mu_s(R_m + R_{ads} + R_{cp} + R_g)} \qquad (5)$$

where R_{ads} is the resistance of solutes adsorbed onto the membrane surface or the pore walls, R_{cp} is the resistance of the concentration polarization boundary layer and R_g is the resistance of a

layer of concentrate at the membrane surface (often referred to as the gel layer). The influence of the different resistances may be established by varying the shear rate.

1.2 CANDIDATE TECHNOLOGIES FOR THE ON-LINE REMOVAL OF VFA FROM THE ANAEROBIC BIOREACTOR

1.2.1 Organic solvent liquid-liquid extraction. This technology requires to put the fermentation broth in contact with solvents that have high selectivity for VFA. There are few organic solvents with this capability, but have been found to be toxic to the microbial population (17).

1.2.2 Adsorption and/or ion-exchange. This technique has received attention for removal of VFA from dilute fermentation broth using basic polymer sorbents. Sorbents containing basic groups provide selective removal of VFA from complex solutions (18). The CELSS anaerobic bioreactor fermentation broth contains a large variety of chemical compounds, such as: proteins, enzymes, inorganics, polysaccharides, ammonia, and many others. Many of these compounds compete for sorbent sites, which require continuous sorbent regeneration and difficult down-stream separation processes.

1.2.3 Dialysis. This is a diffusion mass transfer membrane process. This system provides the capability to remove selectively the compounds of interest by controlling membrane opposite side solution composition (dialyzant). Due to the nature of this process, which is diffusion mass transfer controlled, it takes a long period of time to obtain the desired separation (19)

1.2.4 Electrodialysis. This is a thin ion-exchange membrane compartments between a pair of electrodes separation process. It contains inter-membrane spacers, which are very susceptible to clog by suspended matter. All the ionic compounds, including broth nutrients will also be removed together with the VFA (20, 21).

1.2.5 Pressure driven membrane separation processes. These membrane permselective separation processes are based on particle or molecular size, shape, and other factors affecting the separation. The five major membrane separation processes are: reverse osmosis, for separation of low molecular weight compounds (MW < 1,000); nanofiltration, for molecules with molecular weight (MW) in the range between 1,000 - 10,000; ultrafiltration, for separation of macromolecules with MW in the range of 10,000 to 500,000; and microfiltration for the separation of fine particles in the range of 0.1 to 10 microns. Due to the wide separation spectrum of this technology, many biotechnological separations can be carried out by means of selecting the appropriate membranes and optimum operating conditions.

1.3 PROPOSED TECHNOLOGY

The technique proposed for the on-line removal of VFA from the anaerobic bioreactor fermentation broth is a nanofiltration membrane recycle bioreator (MRB, see Fig. 1). This system consists of a reaction vessel operated as a stirred tank reactor, which is coupled in a semi-closed loop via a suitable pump to a membrane filtration module. The content of the bioreactor is continuously pumped through the membrane module and recycled back to the vessel. The membrane should be chosen to retain the microorganisms, suspended matter, enzymes, and most macromolecules, while minimizing retention of end-products. A MBR using nanofiltration membranes with nominal molecular weight cut-of (MWCO) between 1,000 to 10,000 will greatly remove VFA from the anaerobic bioreactor inedible biomass medium. Due to the high content of suspended matter and large particle sizes of the bioreactor medium is recommended to use tubular membrane configuration systems with large internal diameter. The membrane module selected to accomplish with this requirement was a porous carbon support, internally coated with a ZrO_2-TiO_2 layer, with a length of 0.4 m and an internal diameter of 0.6 cm (Rhône-Poulenc Tech-Sep, Cranbury, New Jersey).

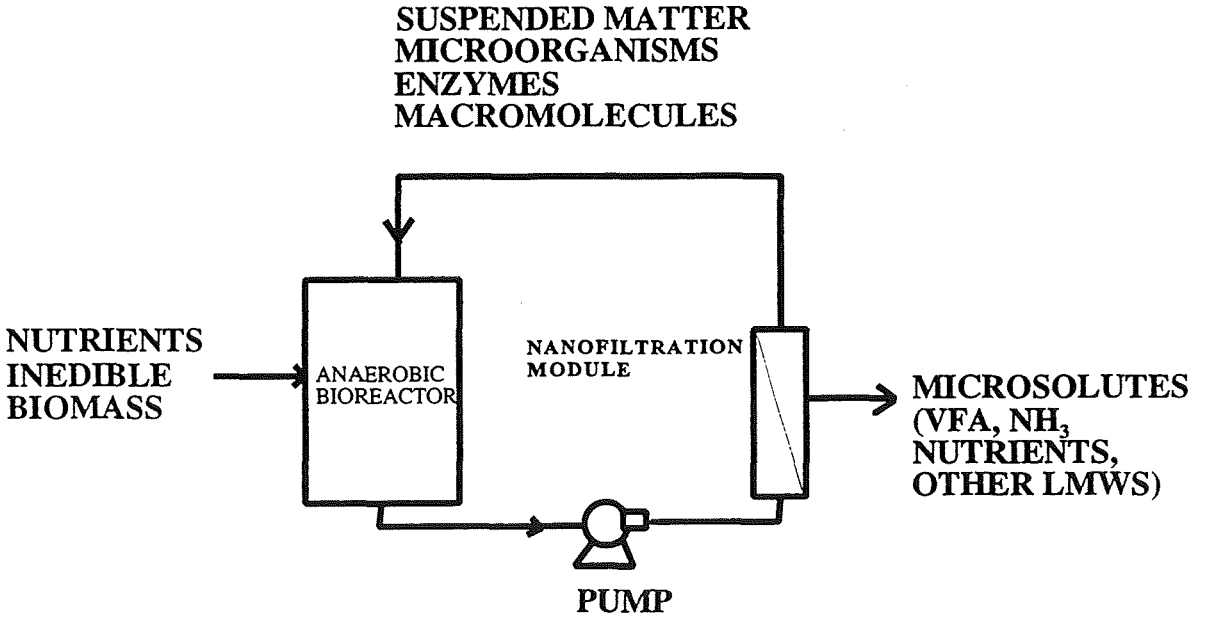


Figure 1. Proposed Nanofiltration Membrane Recycle Bioreactor for On-line Removal of Volatile Fatty Acids from CELSS Anaerobic Bioreactor.

1.4 OBJECTIVE

The main objective of this summer project was to determine the nanofiltration membrane separation process performance variables using a 10,000 MWCO tubular ceramic membrane module before coupling it to the anaerobic bioreactor. The variables to be studied are: the bioreactor slurry permeation characteristics such as; permeate flux as a function of transmembrane pressure, the effects of dissolved and suspended matter content on permeate flux, the effects of the recirculation hydrodynamic conditions on membrane fouling, on-line VFA removal rate and nutrients removal rate.

II. EXPERIMENTAL SET-UP

2.1 EXPERIMENTAL EQUIPMENT

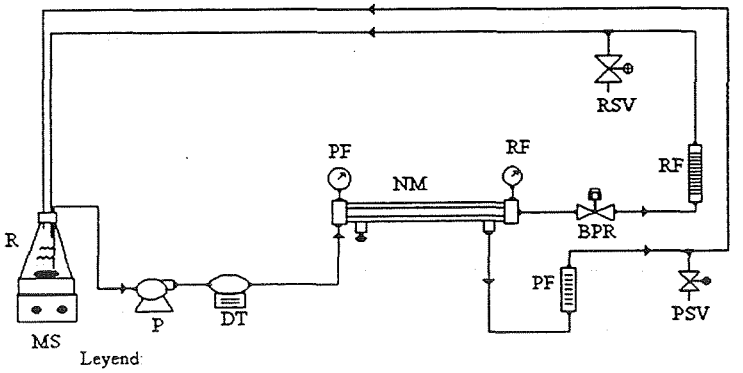
The experimental equipment flow diagram is shown in Fig. 2. The nanofiltration module is constituted of a tubular ceramic membrane composed of a porous carbon support internally coated with a ZrO_2-TiO_2 layer, with a length of 0.4 m and an internal diameter of 0.6 cm (Rhône-Poulenc Tech-Sep, Cranbury, New Jersey). The total membrane filtration surface area is 0.008 m^2 and has a nominal molecular weight of 10,000 Molecular Weight Cut-Off (MWCO).

2.2 ANALYTICAL METHODS

Volatile fatty acids (VFA), such as acetic, propionic, and butyric acid concentrations were determined with a Perkin Elmer Series 4 High Performance Liquid Chromatograph, using a Bio-Rad Aminex HPX-87H column. The mobile phase was a 0.008N H_2SO_4 solution at a flow rate of 1 mL/min and 55°C. Detention was performed with a Perkin Elmer UV Detector at a wavelength of 210 nm. Samples were prepared by mixing 5 mL of the original sample with 1mL of 25%(w/v) metaphosphoric acid to separate out the VFA and then centrifuged for 20 min at 14,000 rpm.

Inorganic analysis of calcium, phosphate, ammonium, and iron concentrations were carried out using EPA methods: EPA 200.7, EPA 365.1, and EPA 350.1. Analyses of total organic carbon (TOC) concentrations were determined using the EPA method: EPA 415.1. Total suspended solids (TSS) concentrations were determined by weighting known samples volume previously dried in a vacuum oven at 75°C and 0.5 bar of vacuum for 24 hours.

Fig. 2. Nanofiltration Experimental Equipment Flow Diagram



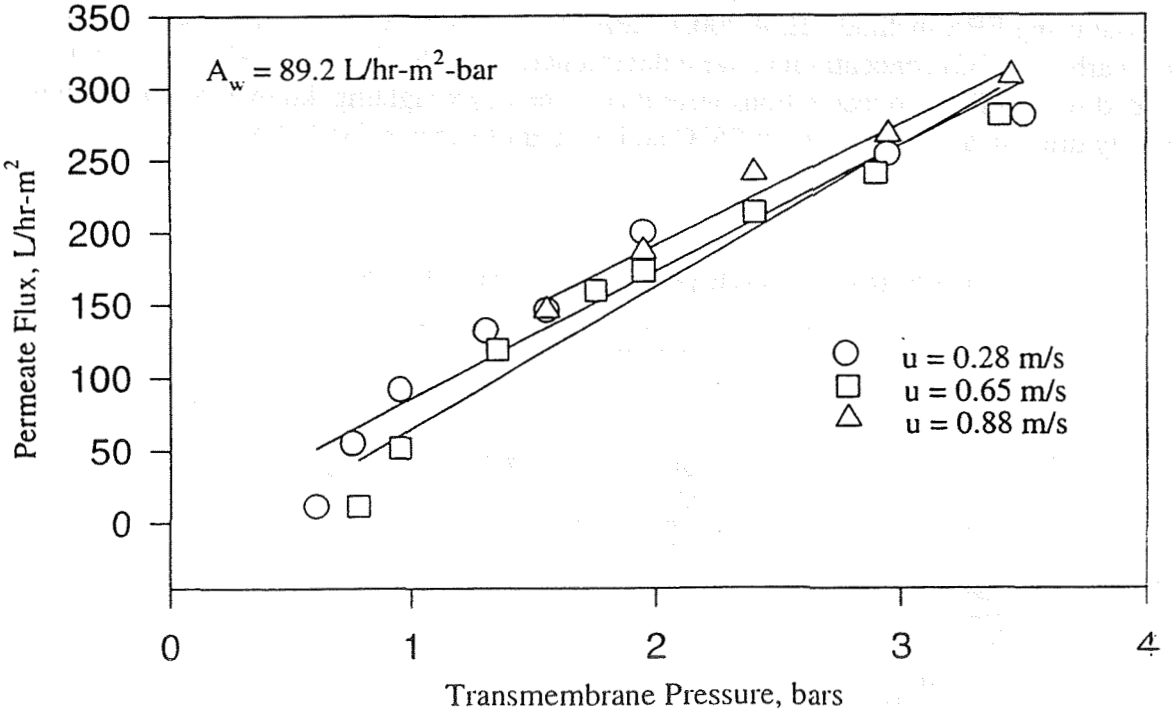
Leyend
R = Reservoir Tank
MS = Magnetic Stirrer
P = Peristaltic Pump
DT = Damping Tank
PF = Feed Pressure Meter
PSV = Permeate Sampling Valve
PR = Retentate Pressure Meter
NM= Ceramic Membrane Module
BPR = Back Pressure Regulator
RSV = Retentate Sampling Valve
RF = Retentate Flowmeter
PF = Permeate Flowmeter

III. RESULTS AND DISCUSSION

3.1 MEMBRANE CHARACTERIZATION

The hydraulic permeability of the membrane was determined before subsequent use in the separation studies. Fig. 3. shows the de-ionized water permeate flux versus applied transmembrane pressure relationship at three different recirculation velocities. The average membrane water permeability constant, A_w, as determined by Eq. 3, is 89.2 L/hr-m^2-bar. It is almost independent on fluid recirculation velocity. The membrane does not have any degree of compressibility under pressure, as indicated there is not deviation from linearity in the profile. After each experimental run using the effluent of the anaerobic bioreactor the membrane was cleaned, following manufacturer procedures for biotechnology fluids. After the cleaning procedure the de-ionized water permeability constant was checked to be a least 95% of the initial value.

Fig. 3. De-ionized Water Permeate Flux vs Transmembrane Pressure at Different Recirculation Velocities



3.2 ANAEROBIC BIOREACTOR EFFLUENT CHARACTERISTICS

The CELSS bench-scale anaerobic bioreactor is a 8 liter working volume reactor. The bioreactor is fed with a previously dried and ground potato inedible plant parts, and is inoculated with culture of rumen bacteria. This microbial population has the enzymatic capability to degrade the cellulose and hemicellulose portion of the biomass. Nitrogen is pumped into the bioreactor so that pressure is high enough to keep oxygen from flowing in, therefore maintaining the anaerobic nature of the system. The bioreactor is operated in a feed and bleed mode of operation, one liter of the bioreactor effluent is harvested each day and replaced with one liter of de-ionized water and 50 g of inedible biomass.

To carry out each nanofiltration experimental run 3 liters of harvested bioreactor effluent were collected. The bioreactor effluent composition of total suspended solids (TSS), volatile suspended solids (VSS), total organic carbon (TOC), volatile fatty acids, and some inorganic nutrients is given in Table 1. The bioreactor effluent V.A. composition is about 84% acetic acid, 8.6% propionic acid, and 7.4% butyric acid (molar percent). As can be observed from Table 1., it contains a large amount of suspended matter.

The bioreactor effluent also contained some large particles, which clogged the back pressure regulator connected at the outlet of the nanofiltration module. To apply a given transmembrane pressure, it is required to close this valve slowly, reducing the available fluid flowing opening area. Some of the larger particles contained in the bioreactor effluent clogged it and the pressure rise-up drastically, producing unstable operating conditions. To solve this problem the bioreactor effluent was filtered with an 0.04 x 0.04 mm opening stainless steel wire mesh. The filtration step reduced the total suspended solids concentration from 36,972 to 24,648 mg/L (33% reduction on TSS concentration). The total suspended solids concentration of the bioreactor effluent before and after the filtration step changed from one to another harvest. The filtered effluent total suspended solids concentration varied from 26,500 to 23,000 mg/L. In practice this filtration step can be done by installing an on-line filter probe to the bioreactor when it will be coupled to the nanofiltration module.

3.3 EFFECT OF APPLIED TRANSMEMBRANE PRESSURE AND FLUID RECIRCULATION VELOCITY ON TOTAL FLUID PERMEATE FLUX, VFA AND NUTRIENT REMOVAL RATES.

Several experiments were carried out to establish the best nanofiltration operating conditions. At fluid recirculation velocities lower than 0.5 m/s, the pump was not able to keep a constant flow rate operation. This phenomenon is mostly produced due to the high content of suspended matter and high viscosity of the anaerobic bioreactor effluent, where the pressure forces are not enough to overimpose viscous forces. The operation at fluid velocities higher than 1.2 m/s generated a persistent foam which produced pump cavitation. Therefore, the most stable fluid recirculation velocities were found to be between 0.6 to 1.0 m/s.

Table 1. Anaerobic Bioreactor Effluent Average Composition.

Analysis	Composition
TSS, mg/L	36,972
VSS, mg/L	25,819 (70% of TSS)
TOC, mg/L	4215
Acetic Acid, mMole	147.4
Propionic Acid, mMole	15
Butyric Acid, mMole	13.1
PO_4-P, mg/L	233.3
NH_4-N, mg/L	532
Calcium, mg/L	75.6
Iron, mg/L	12.1

The permeate flux as a function of applied transmembrane pressure at three different velocities is shown in Fig. 4. As can be observed, the permeate flux increases with applied transmembrane pressure and fluid recirculation velocity. At applied pressure lower than 1.5 bar the permeate flux is directly proportional to the transmembrane pressure. At applied pressure higher than 1.5 bar the permeate flux is not more linearly dependent on pressure. This asymptotic flux-pressure relationship observed at high applied pressure is due to the effects of the concentration polarization and fouling over the membrane surface. As shown in Fig. 5, the permeate flux is also directly proportional to the fluid velocity. This means that increasing the tangential fluid velocity increases the boundary layer mass transfer coefficient, thus minimizing the concentration polarization layer and therefore produce higher permeate flux. For example, the operation at ΔP_m of 1.5 bar and u = 0.65 m/s the permeate flux is close to 15 L/hr-m^2, while at the same applied pressure but at velocity of 1.0 m/s, the flux is 30 L/hr-m^2. The filtration rate is increased by 100% when the fluid recirculation velocity is increased by 67 %.

Figures 6, 7 and 8 show the effect of applied transmembrane pressure and fluid recirculation velocity on VFA removal rate. As can be observed in these Figures, the VFA removal rates have a similar permeate flow pattern as the bulk fluid filtration rate. The VFA removal rate is directly proportional to the bulk permeate flux, because these acids are practically not retained by the membrane. It was found that the concentration of these acids in the feed stream to the membrane module was very similar to those in the permeate stream. The rate of

Fig. 4. Bioreactor Effluent Permeate Flux vs Transmembrane Pressure at Different Fluid Recirculation Velocities

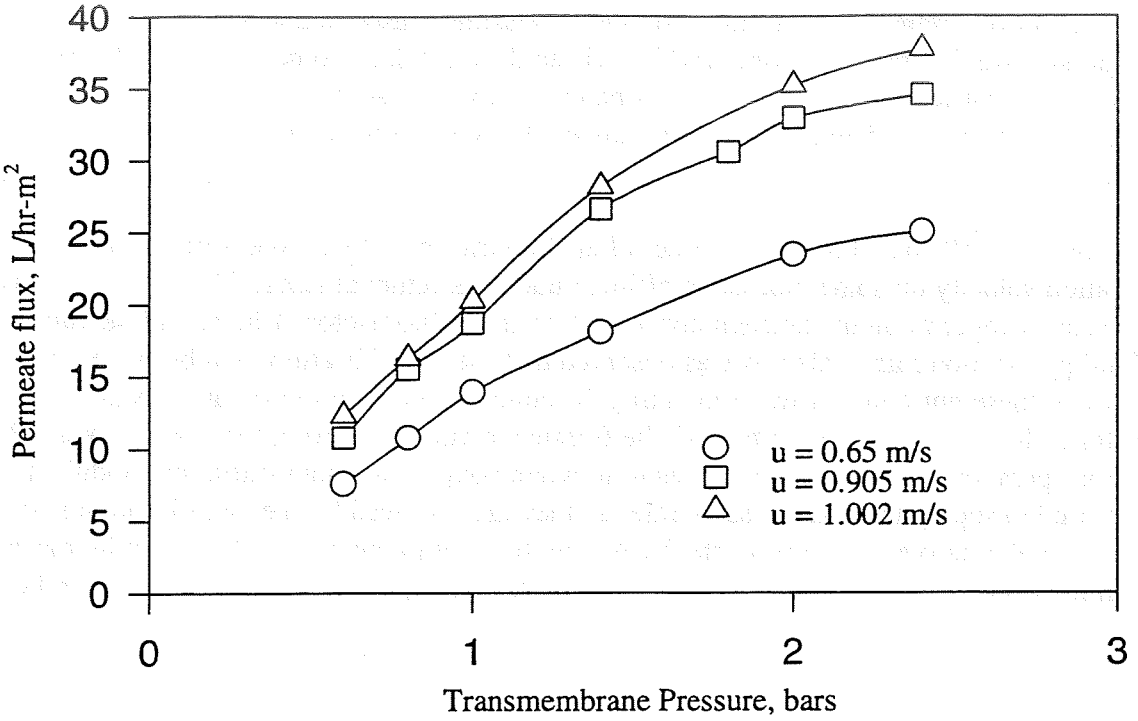
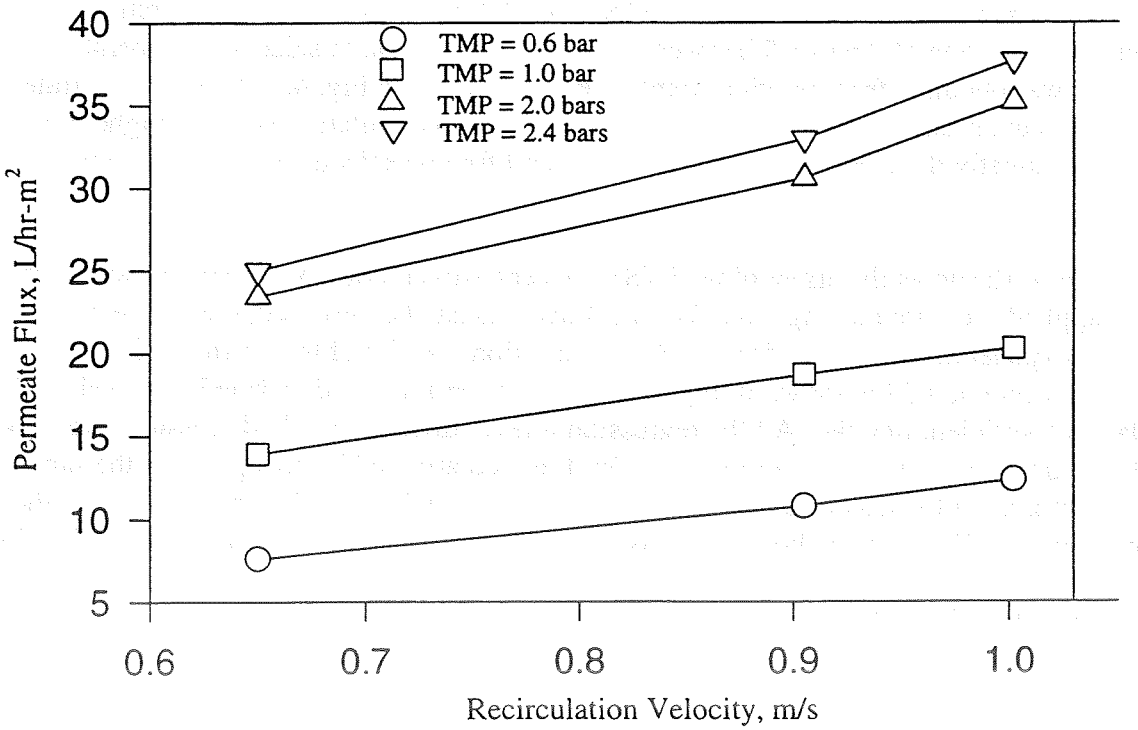


Fig. 5. Bioreactor Effluent Permeate Flux vs Fluid Recirculation Velocity at Different Applied Transmembrane Pressure



removal of acetic acid is higher than both propionic and butyric acids together. The rate of acids removal is proportional to their concentration in the feed stream. As can be observed in Table 1., the VFA is mostly composed of acetic acid. Other possible reason is that acetic acid has a molecular size smaller than propionic and butyric acid, which let it to pass through the membrane pores with less restriction . It is possible to obtain low total VFA removal rate as 60 g/hr-m^2 at ΔP_m = 0.6 bar and u = 0.65 m/s, and high removal rate as 340 g/hr-m^2 at ΔP_m = 2.4 bar and u = 1.0 m/s.

Figures 9, 10 and 11 show the effect of applied transmembrane pressure and fluid recirculation velocity on some bioreactor effluent nutrients removal rates. The rate of removal of these nutrients depends on the nutrient concentration in the bioreactor effluent and the rate of total fluid permeation rate. Also, is was observed that the nanofiltration membrane did not retain any of these nutrients. It was found that the concentrations of these nutrients in the bioreactor effluent were similar to that of the filtrate stream. Therefore, for a proper anaerobic bioreactor operation to avoid nutrients wash-out when coupled to a nanofiltration module it is then required to supply them at the same rates as they are removed by the nanofiltration system. One possible alternative is to remove the VFA from the membrane permeate stream by means of an anion ion exchange process and then recycled back these nutrients to the bioreactor and/or plants.

3.4 EFFECT OF TOTAL SUSPENDED SOLIDS CONCENTRATION ON TOTAL PERMEATE FLUX, VFA AND NUTRIENT REMOVAL RATES.

To study the effect of total suspended solids (TSS) concentration on nanofiltration process performance, a high recirculation velocity of 0.9 m/s, and applied transmembrane pressures in the range of 0.6 to 1.5 bar were selected. The reason to select these operating conditions was obtained from previous experiments as shown in Fig. 4., which demonstrated that the operation at pressure lower than 1.5 bar and high fluid recirculation velocity higher than 0.6 m/s reduced greatly the concentration polarization and fouling effects over the membrane surface.

Figure 12 shows the effect of feed TSS concentration on the membrane filtration rate for process applied pressure ranging from 0.6 to 1.4 bar. As can be observed, the permeate flux decreased exponentially with the feed TSS concentration. A plot of the permeate flux versus the natural logarithm of TSS is shown in Fig. 13. All the experimental data fitted very well with a correlation coefficient of 0.99. All the regression lines at different applied pressures are almost parallel, which means that the operation at a fluid recirculation of 0.9 m/s produce the same degree of mixing or turbulence over the membrane boundary layer and is independent of the applied pressure. The average boundary layer mass transfer coefficient obtained from the slopes of the regression lines and using Equation 1., is 5.98 x 10^{-6} m/s. It is found to be almost independent of applied TMP.

Fig. 6. Volatile Fatty Acids Removal Rate from Bioreactor Effluent at a Recirculation Velocity of 0.65 m/s

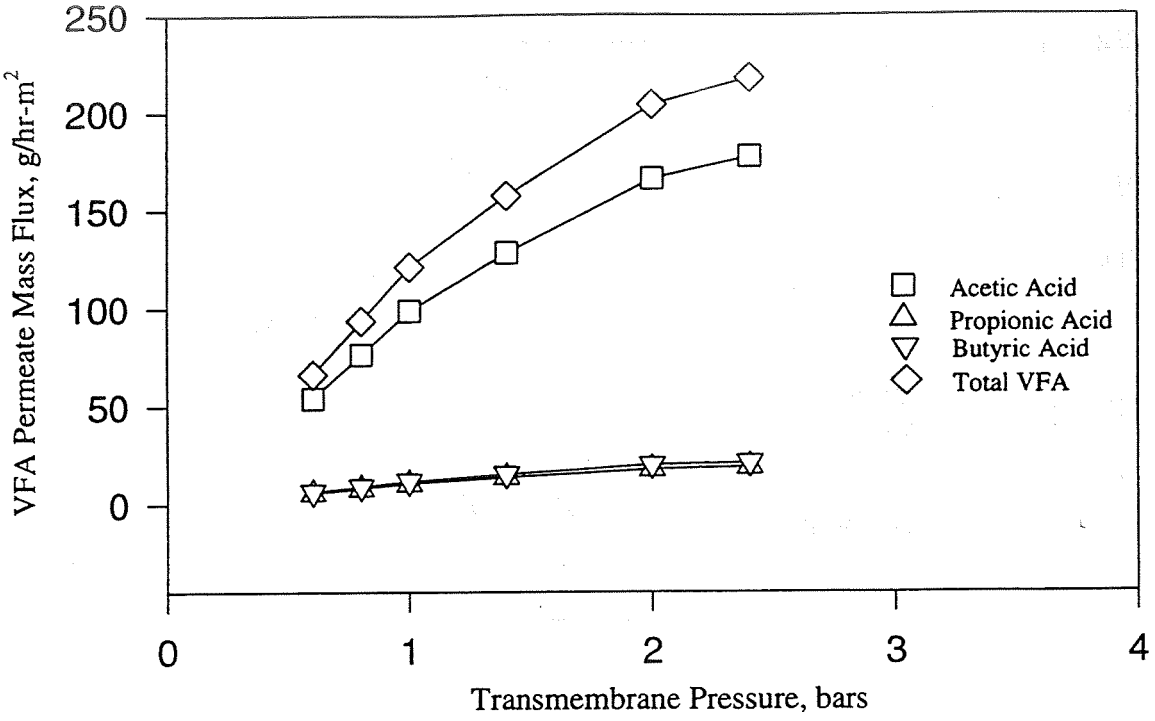


Fig. 7. Volatile Fatty Acid Removal Rate from Bioreactor Effluent at Recirculation Velocity of 0.90 m/s

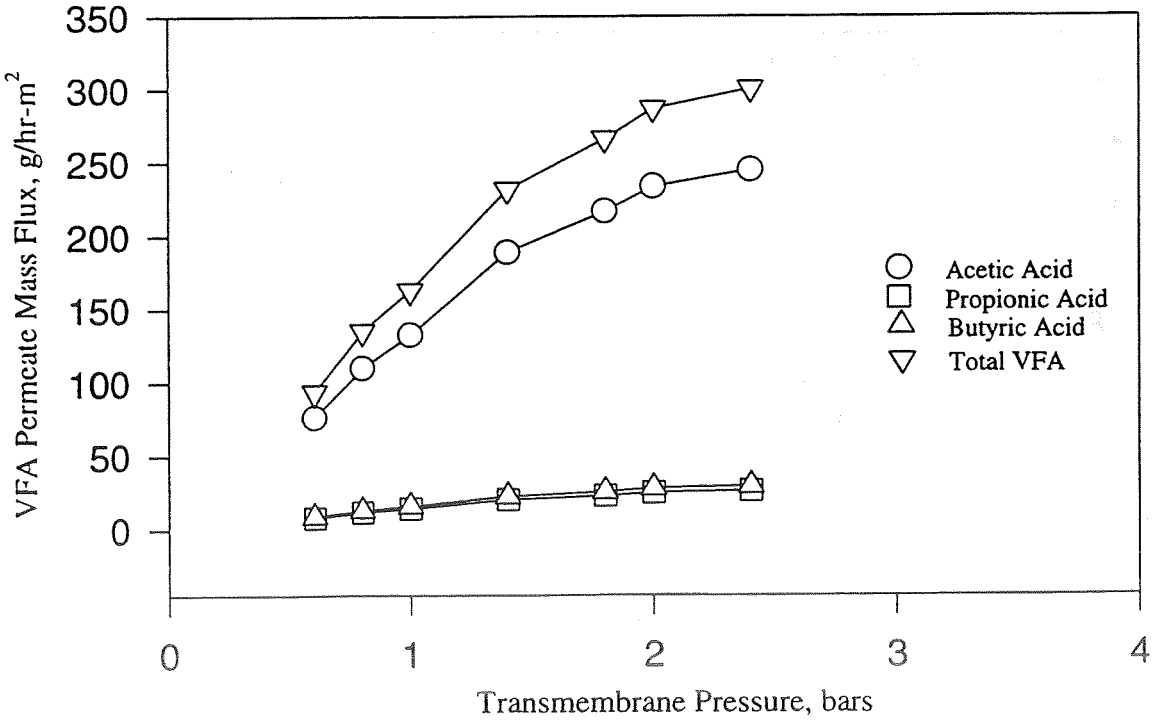


Fig. 8. Volatile Fatty Acids Removal Rate from Bioreactor Effluent At a Recirculation Velocity of 1.0 m/s

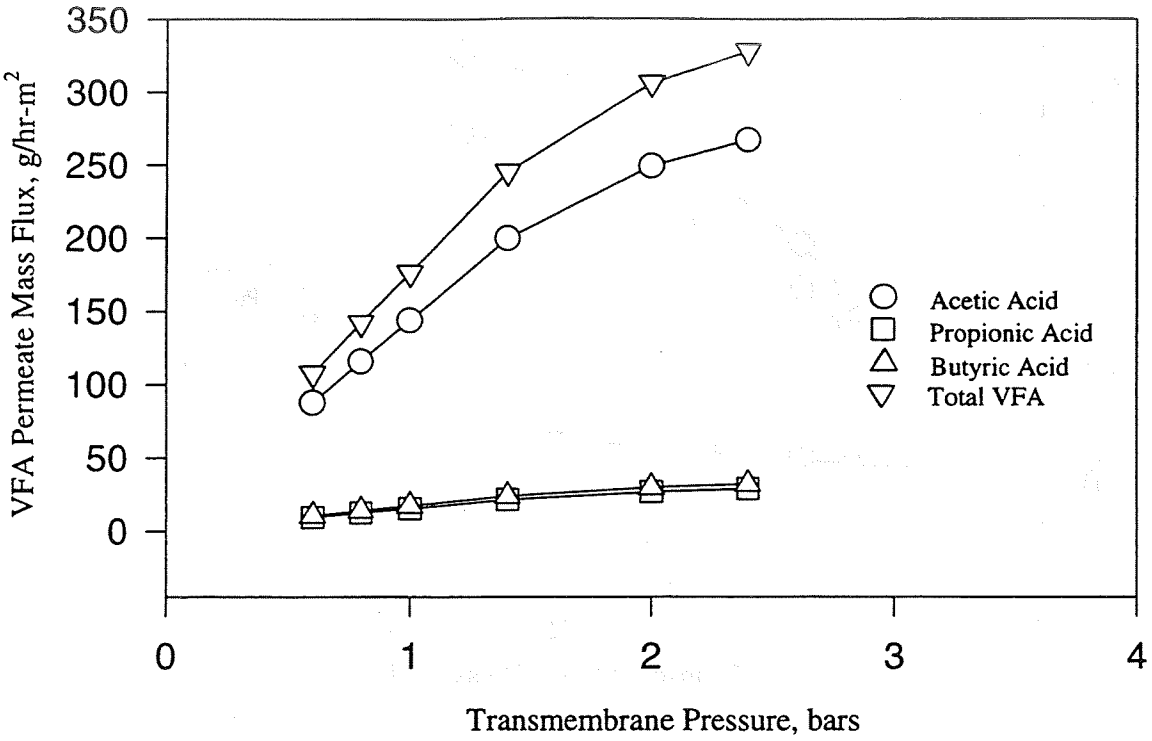


Fig. 9. Nutrients Removal Rate from Bioreactor Effluent at a Recirculation Velocity of 0.65 m/s

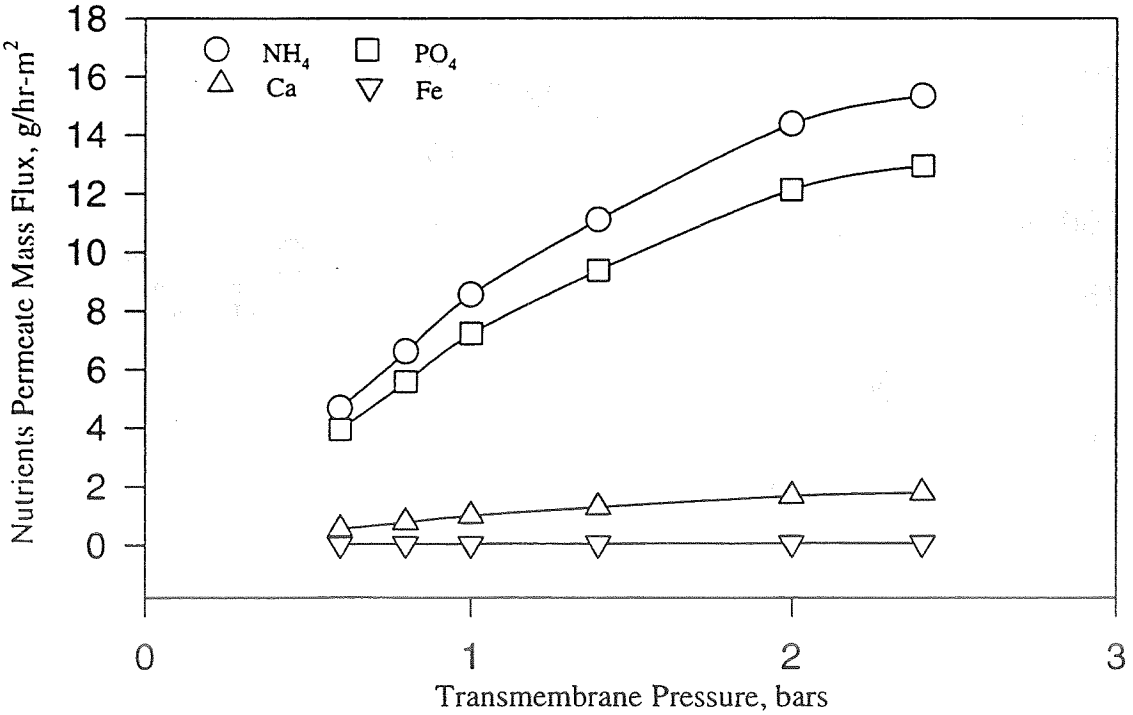


Fig. 10. Nutrients Removal Rate from Bioreactor Effluent at a Recirculation Velocity of 0.90 m/s

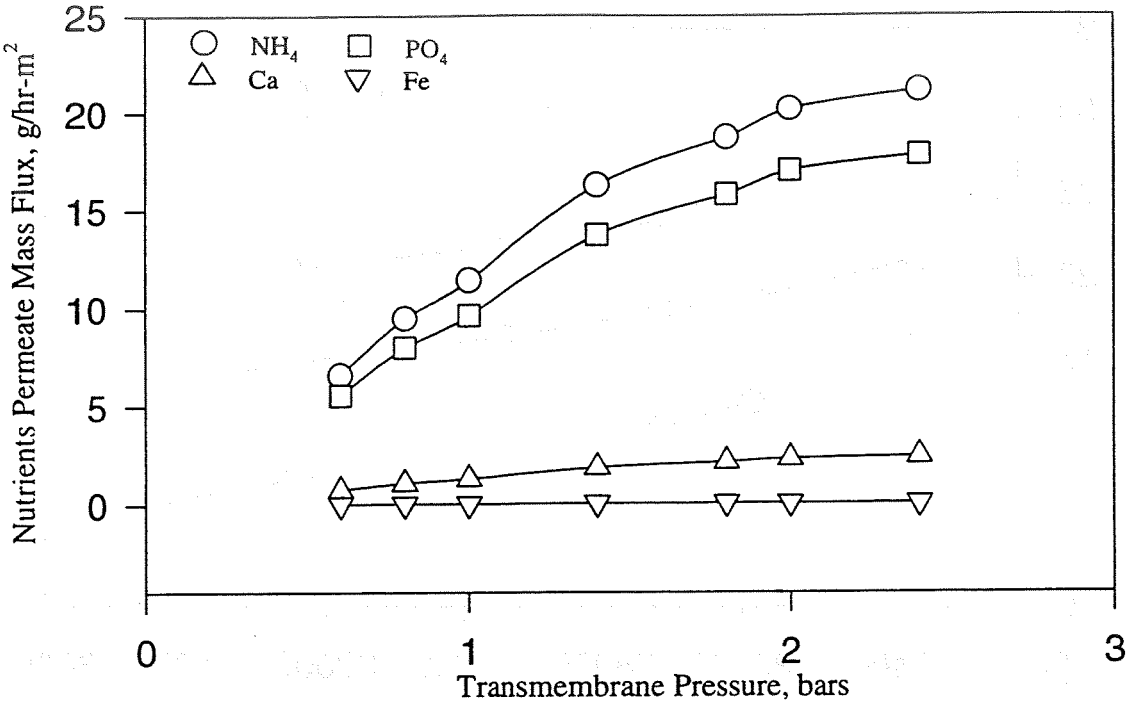


Fig. 11. Nutrients Removal Rate from Bioreactor Effluent at a Recirculation Velocity of 1.0 m/s

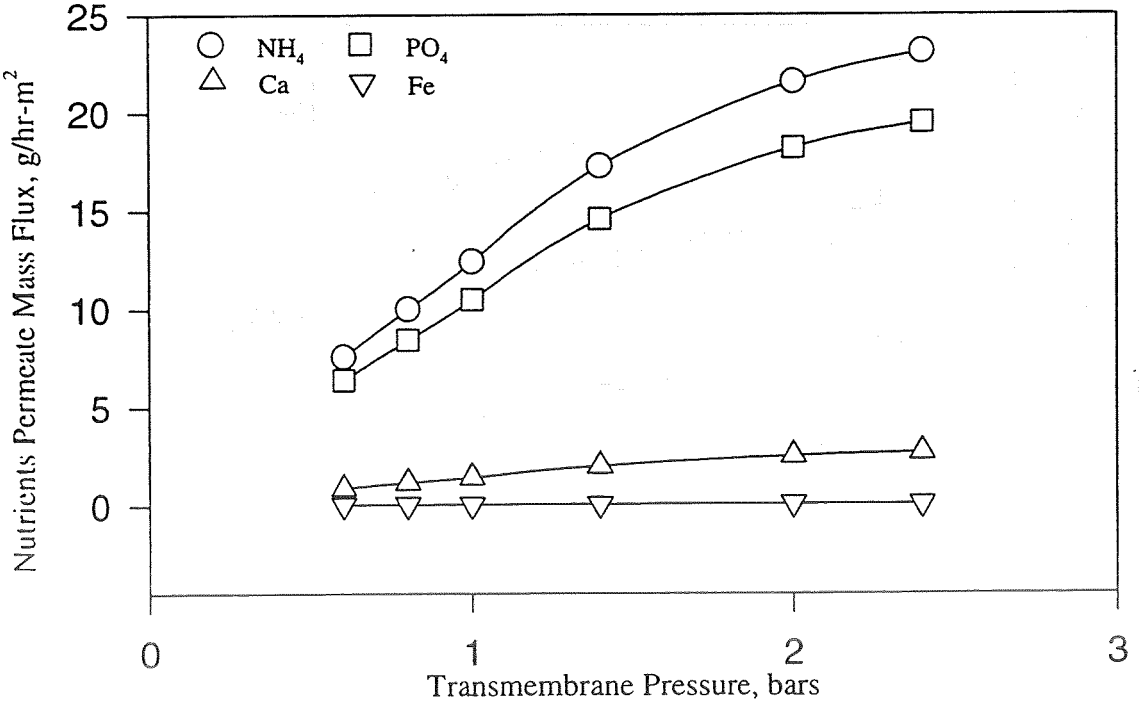


Fig. 12. Effect of Bioreactor Effluent Total Suspended Solids Concentration on Permeate Flux at U = 0.9 m/s

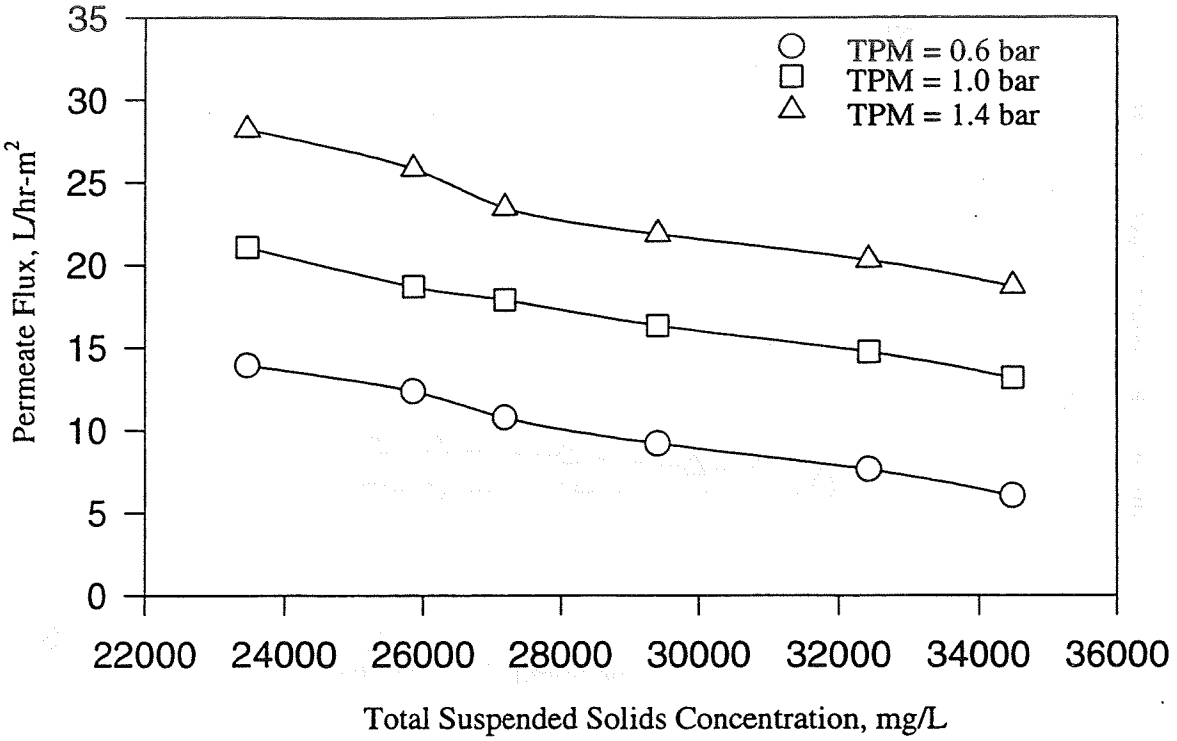
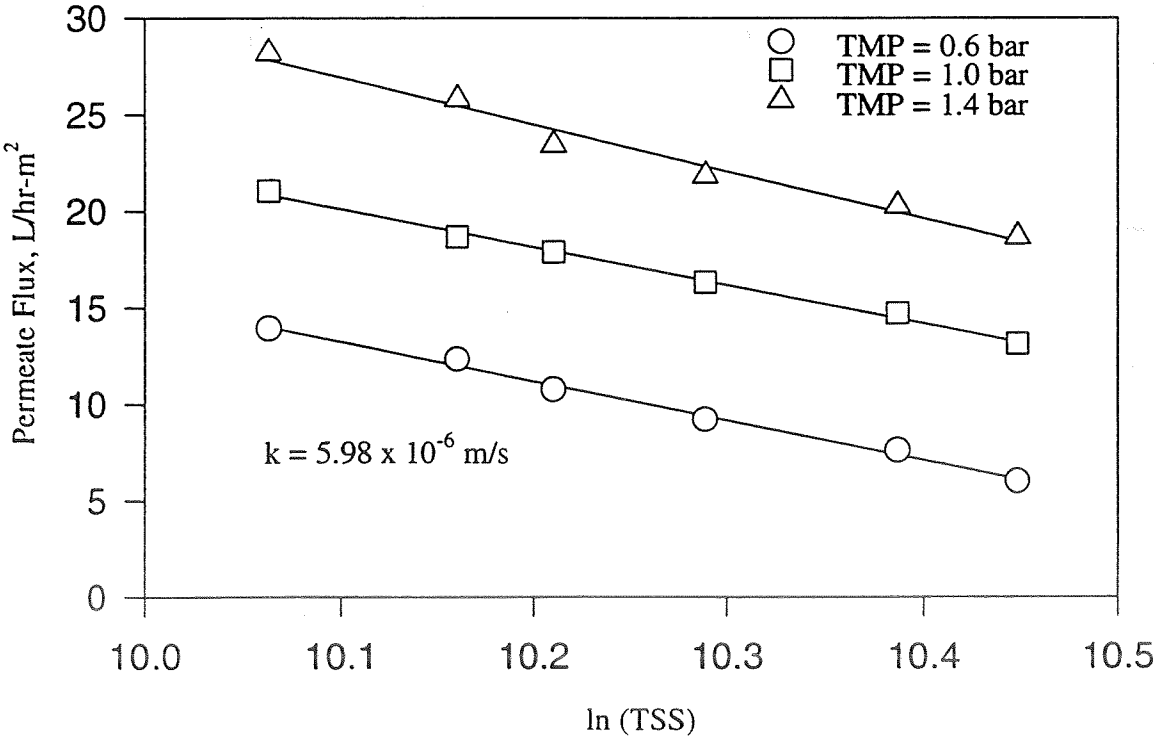


Fig. 13. Effect of Bioreactor Effluent Total Suspended Solids Concentration on Permeate Flux at U = 0.9 m/s (Semi-log Plot)



As noted in Equation 3, 4, and 5 the permeate flux is inversely proportional to solution viscosity. Any increase in the feed concentration will also increase the medium viscosity and therefore, affect inversely the membrane filtrate rate. Also, as the feed TSS concentration increases, there is more resistance of the fluid moving from the solution bulk to the membrane surface due to polarization and gel layers over the membrane surface.

The feed TSS concentration was changed from 23,466 to 34,880 mg/L, which represent a 47% increase in TSS concentration. For this increase in TSS concentration, the permeate flux was reduced from 13.9 to 6.0 L/hr-m^2 (59% reduction) at ΔP_m = 0.6 bar, from 21.1 to 13.2 L/hr-m^2 (37% reduction) at ΔP_m = 1.0 bar, and from 28.2 to 18.7 L/hr-m^2 at ΔP_m = 1.4 bar. As observed in Fig. 14, the permeate flux is directly proportional to the applied pressure for all the bioreactor effluent feed TSS concentrations. The operation at u = 0.9 m/s and ΔP_m < 1.5 bar keep the filtration operation with minimum effects of polarization and fouling.

Figures 15, 16, and 17 show the effect of TSS concentration on VFA removal rates operating at u = 0.9 m/s and applied pressure ranging from 0.6 to 1.4 bar. For and increase of feed TSS concentration from 23,466 to 34,880 mg/L, the total VFA removal rate was reduced from 108.8 to 48.0 g/hr-m^2 (56% reduction) at ΔP_m = 0.6 bar, from 164.4 to 104.8 g/hr-m^2 (36% reduction) at ΔP_m = 1.0 bar, and from 220 to 148 g/hr-m^2 (32% reduction) at ΔP_m = 1.4 bar. This means that the total VFA removal rate is directly proportional to the applied pressure, but is inversely proportional to the feed TSS concentration. As observed in these Figures, the removal rate of acetic acid is much higher than the removal rates of both propionic and butyric acids, but the total reduction percent in removal rates due to the increase in TSS concentration are almost the same for the acids.

Figures 18, 19, and 20 show the effect of TSS concentration on some bioreactor effluent nutrients removal rates. The nutrients removal rates were found to be directly proportional to the applied pressure, total filtrate rate, nutrient concentration in the bioreactor effluent (see Table 1), and inversely proportional to the TSS concentration.

3.5 EFFECT OF PROCESS OPERATING TIME ON FLUID PERMEATE FLUX.

Figure 21 shows the effect of the nanofiltration membrane separation process operating time on permeate flux at u = 0.9 m/s and applied pressure ranging from 0.6 to 1.5 bar. The concentration of feed TSS was kept constant at 23,780 mg/L during all the transient period. As can be observed, the permeate flux declines rapidly during the first 5 to 8 hours, and then levels off with a diminishing rate of flux decay . This occurred as consequence of the accumulation of suspended and some dissolved matter over the membrane surface which no longer participates in the mass-transport to or away from the membrane. A cake or gel layer may be formed which constitutes a considerable hydraulic resistance. At operating time higher than 8 hours the flux-time relationship reaches asymptotically steady-state operating conditions. This means that the

Fig. 14. Effect of Applied Transmembrane Pressure on Permeate Flux at Different TSS Concentration at U = 0.9 m/s

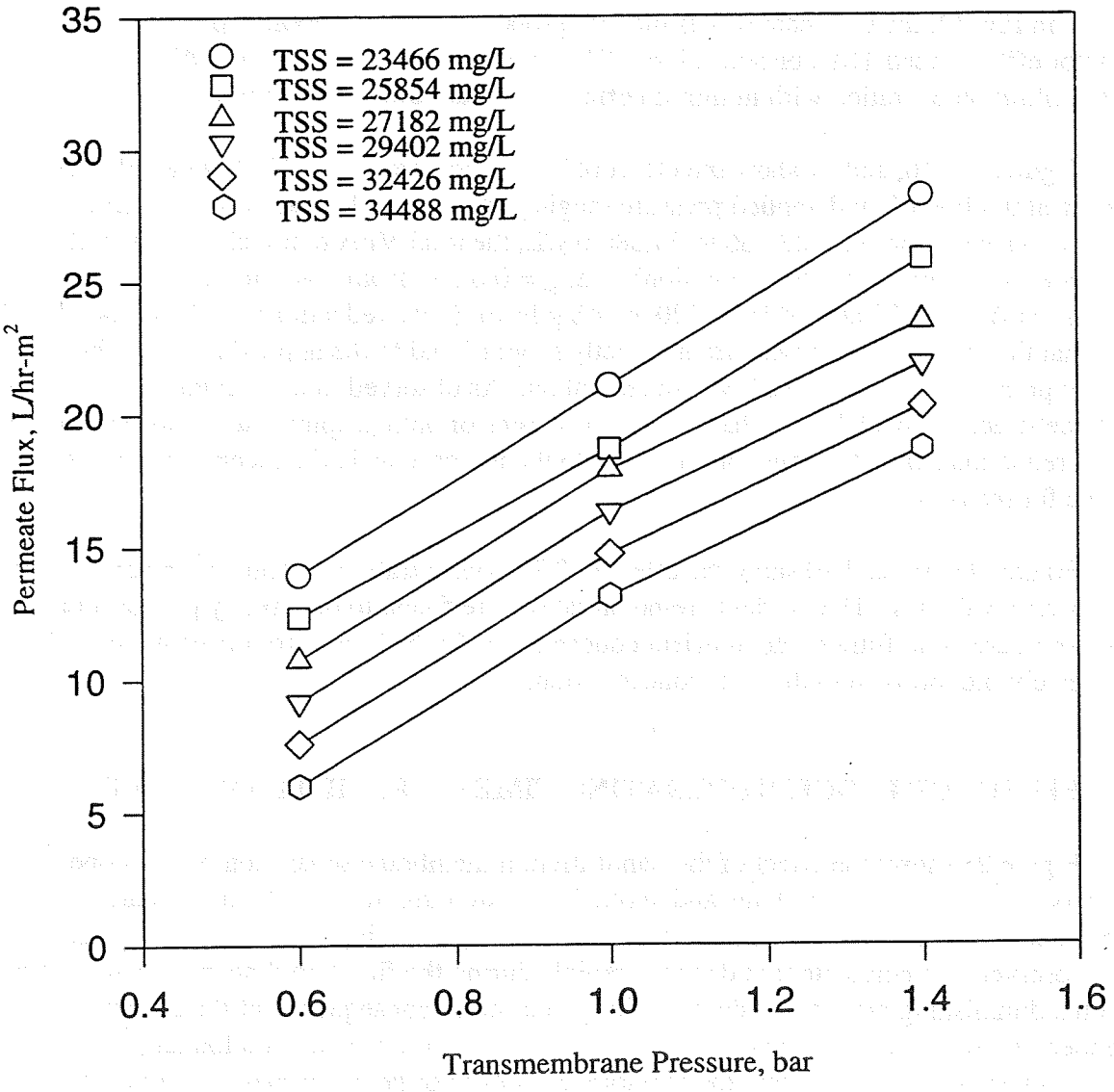


Fig. 15. Effect of Bioreactor Broth Total Suspended Solids Concentration on VFA Removal Rate at TMP = 0.6 bar and U= 0.9 m/s

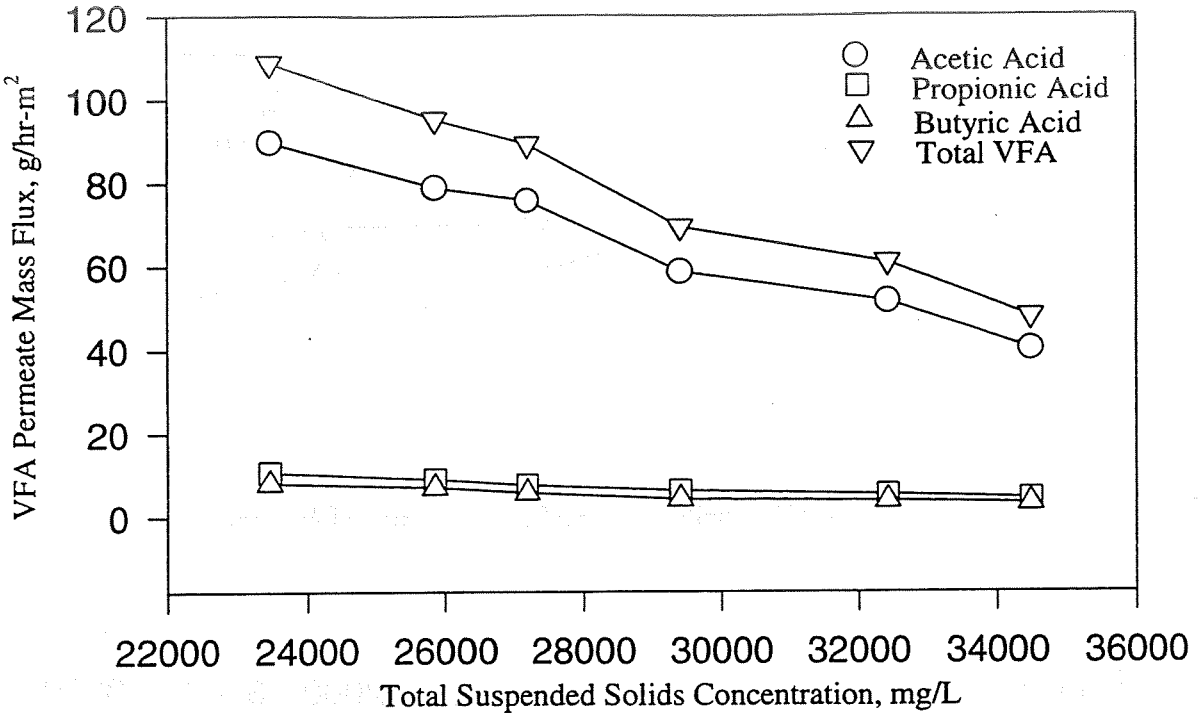


Fig. 16. Effect of Bioreactor Broth Total Suspended Solids Concentration on VFA Removal Rate at TMP = 1.0 bar and U = 0.9 m/s

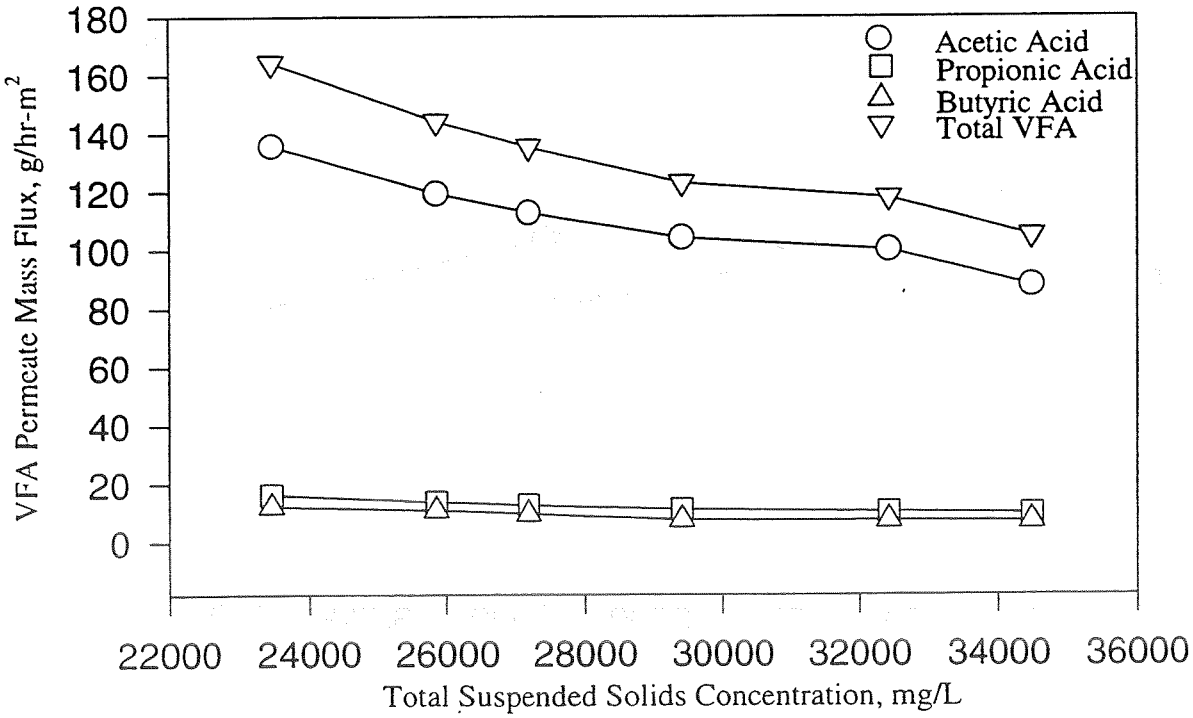


Fig. 17. Effect of Bioreactor Broth Total Suspended Solids Concentration on VFA Removal Rate at TMP = 1.4 bar and U = 0.9 m/s

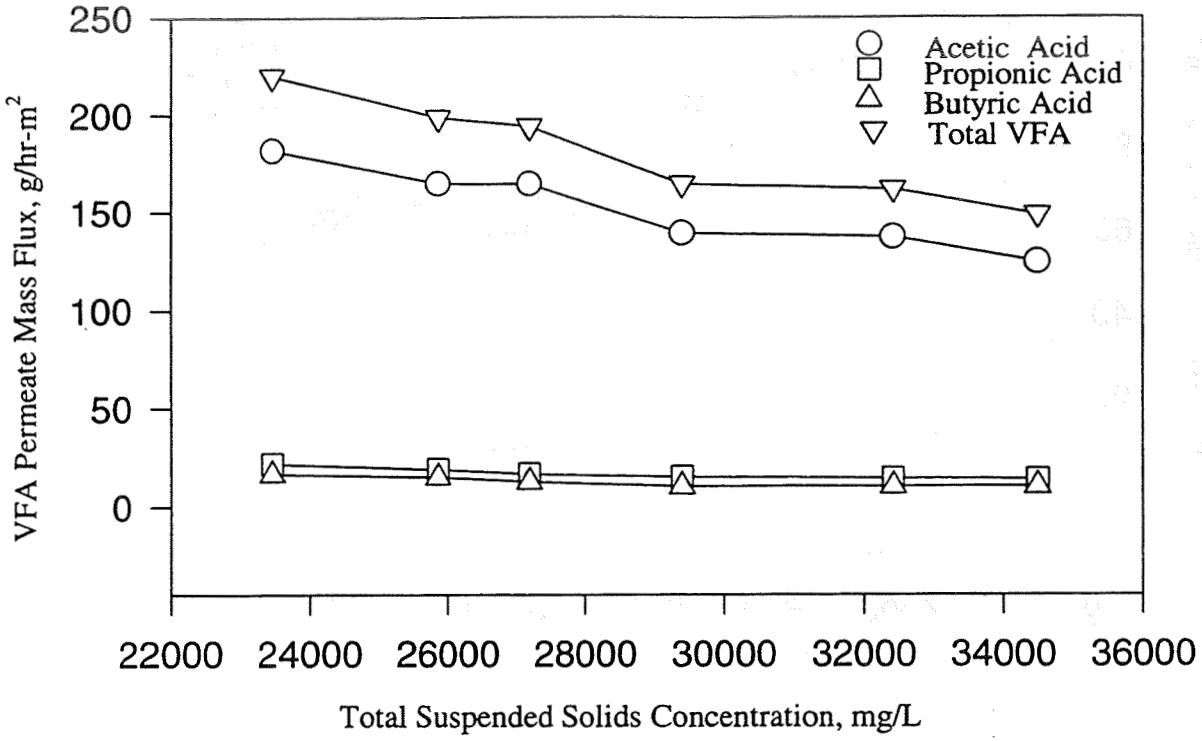


Fig. 18. Effect of Total Suspended Solids Concentration on Nutrients Removal Rate at TMP = 0.6 bar and U = 0.9 m/s

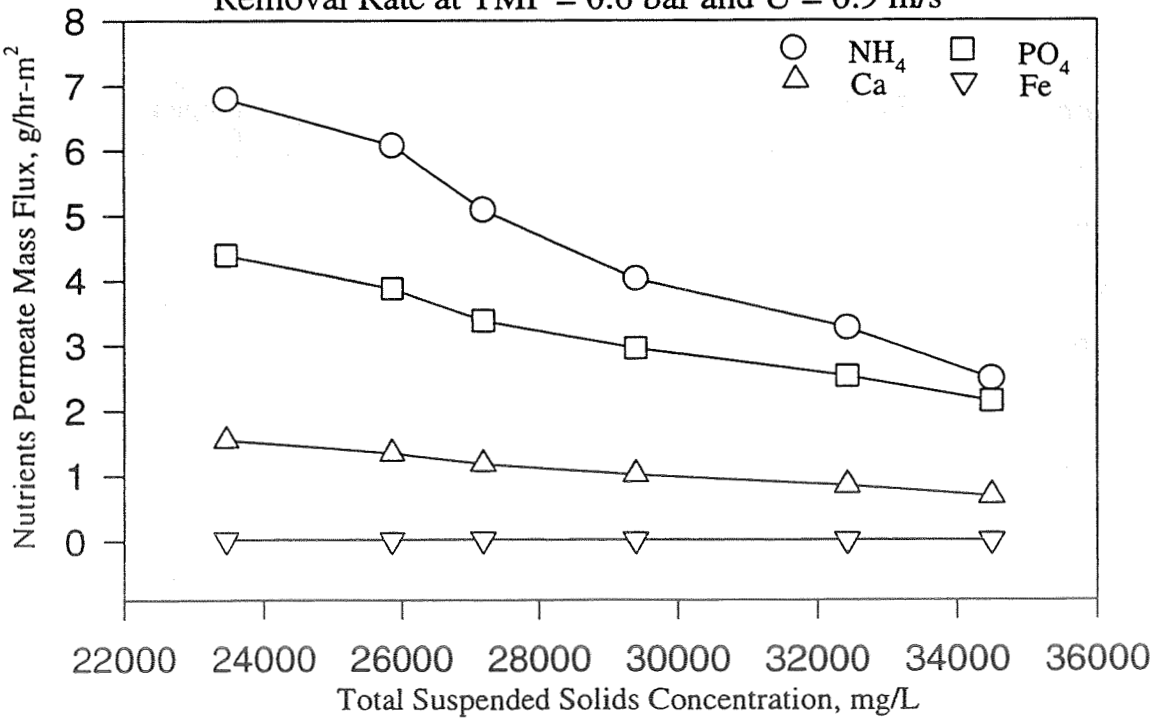


Fig. 19. Effect of Total Suspended Solids Concentration on Nutrients Removal Rate at TMP = 1.0 bar and U = 0.9 m/s

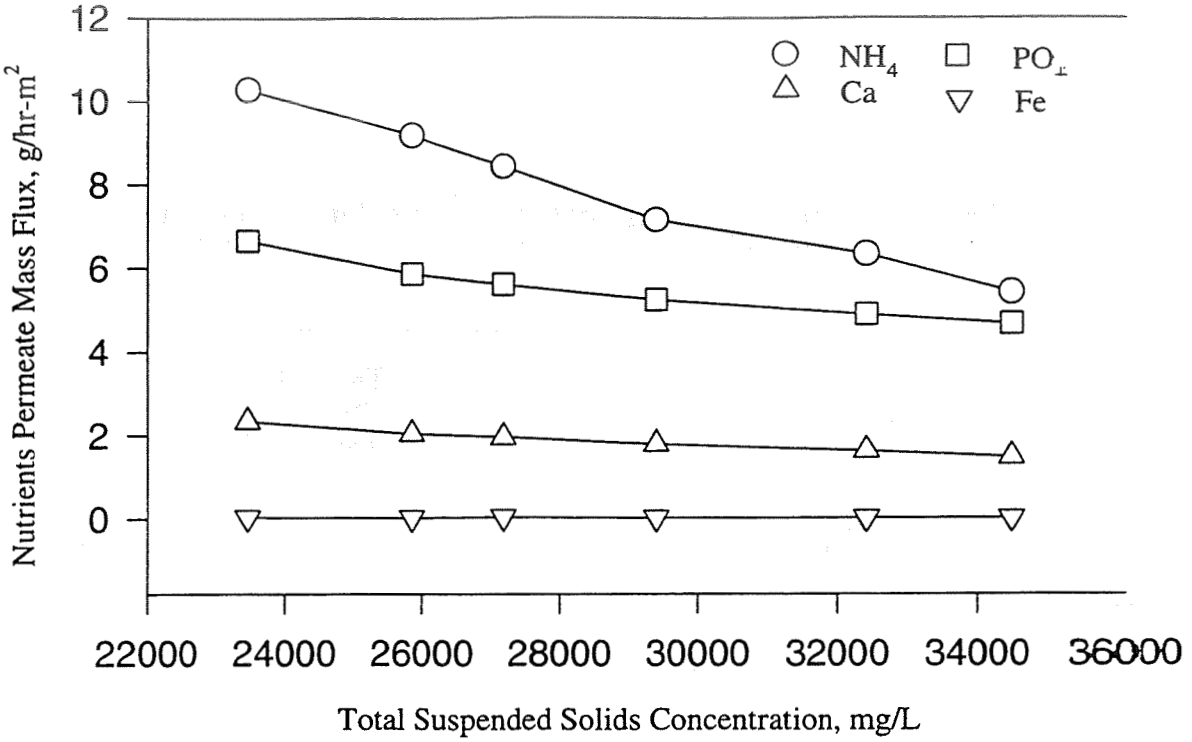


Fig. 20. Effect of Total Suspended Solids Concentration on Nutrients Removal Rate at TMP = 1.4 and U = 0.9 m/s

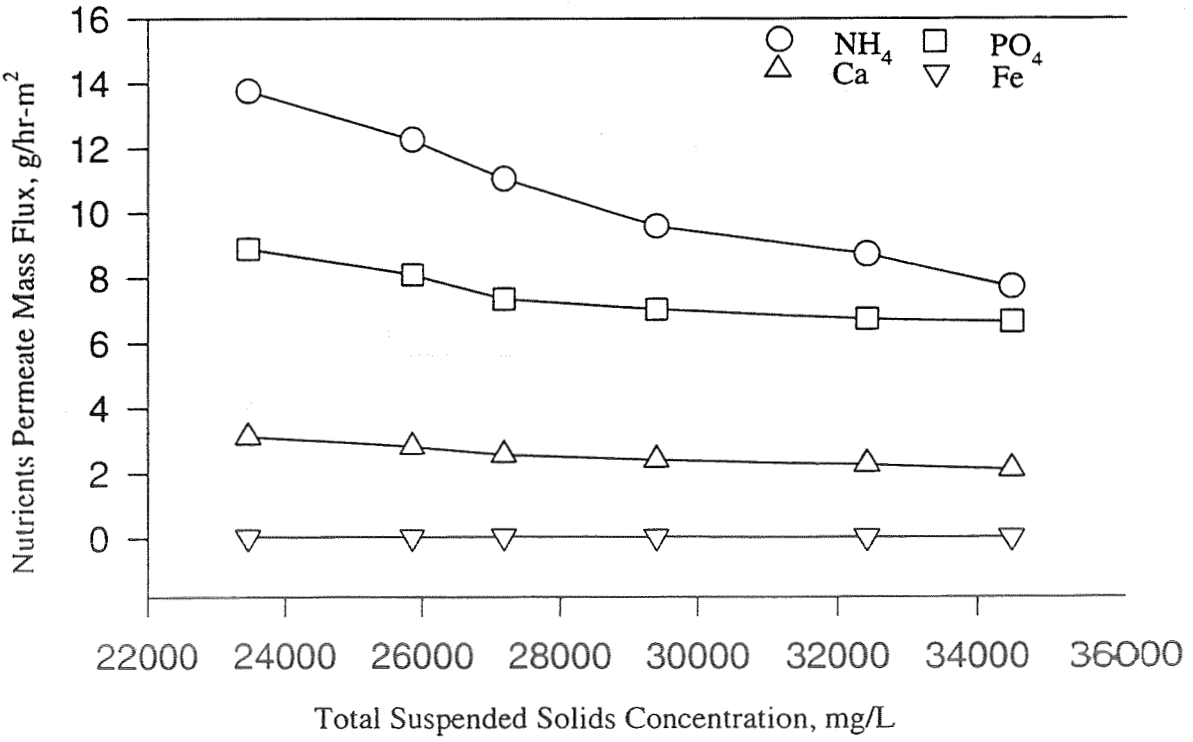
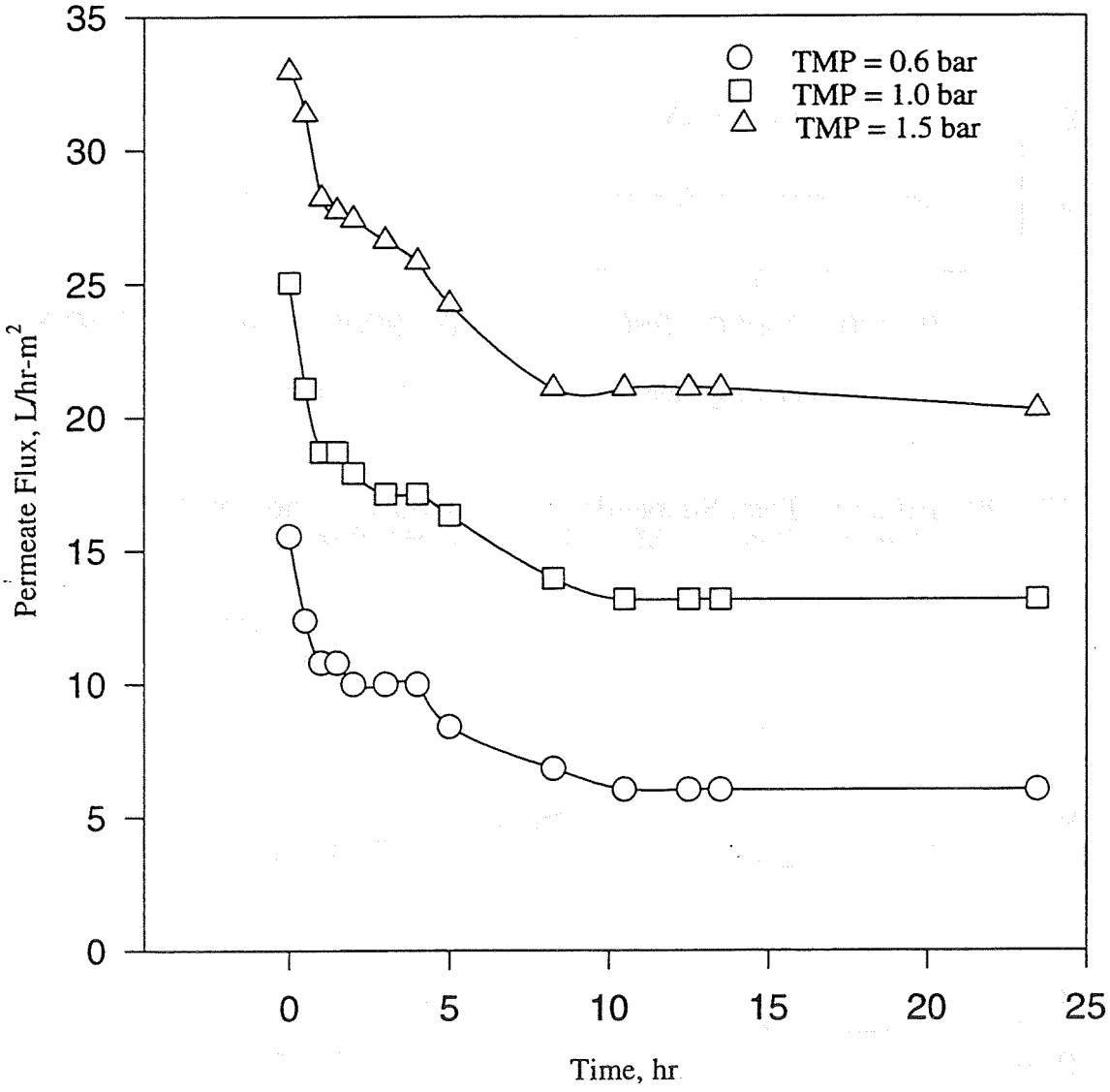


Fig. 21. Variation of Permeate Flux with Time at Different Applied Transmembrane Pressure at U = 0.9 m/s



gel layer thickness is highly susceptible to the shear rate produced by the fluid hydrodynamic conditions (16). The layer thickness reached a finite value which is controlled by the fluid shear rate, and any additional suspended matter deposited over this layer is swept away by the cross-flow fluid.

During the flux-time decline period (0 to 10 hours) the flux decreased from 15.5 to 6.0 L/hr-m^2 (61% flux reduction) at ΔP_m = 0.6 bar, from 25 to 13 L/hr-m^2 (38% flux reduction) at ΔP_m = 1.0 bar, and from 33 to 21 L/hr-m^2 (36% flux reduction) at ΔP_m = 1.5 bar. One practical way to restore the permeate flux is by means of high cross-flow velocities, which produce higher shear rate conditions to sweep deposited matter away, also by interrupting the permeate flux. As the permeate flux is interrupted the cake is not longer forced against the membrane surface by the permeate flux and may flow laterally over the membrane surface and is discharged with the retentate stream.

IV. CONCLUSIONS

Based on the results obtained using a nanofiltration tubular ceramic membrane module to study the anaerobic bioreactor permeation characteristics, the following conclusions were reached:

The bioreactor effluent permeate flux is directly proportional to the fluid recirculation velocity in the range between 0.6 to 1.0 m/s and to the applied transmembrane pressure when this is lower than 1.5 bar. The filtrate rate flux is significantly affected by polarization and fouling effects at pressure higher than 1.5 bar. The total permeate flux decreases exponentially with the increasing of the bioreactor effluent total suspended solids concentration. The permeate flux declines rapidly during the first 5 to 8 hours and then levels off with a diminishing rate of decay. This occurred as result of the accumulation of suspended matter over the membrane surface which no longer participates in the mass-transport to or away from the membrane. One practical way to restore the permeate flux is by means of high cross-flow velocities, which produce high shear rate conditions to sweep deposited matter away, also by means of interrupting the permeate flux.

The volatile fatty acids removal rate from the bioreactor effluent is proportional to the applied pressure, fluid recirculation velocity, acid concentration in the fermentation broth, and permeate flux, but inversely proportional to the total suspended solids concentration. Fermentation medium nutrient, such as ammonium, phosphate, calcium, and iron are also removed in the permeate stream at rates proportional to nutrient concentration in bioreactor medium, total permeate flux, and inversely proportional to TSS concentration.

The nanofiltration separation using ZrO_2-TiO_2 membrane over a carbon support with large internal diameter enough to avoid system clogging proved to be an acceptable technique for the on-line removal of VFA from the CELSS anaerobic bioreactor. Due to bioreactor effluent's nutrients are also removed simultaneously with VFA when coupling it to the bioreactor, it is highly recommended to supply them at the same rate as they are removed. One possible alternative to recover permeate stream nutrients is to remove the VFA by mean of basic ion-exchange resins and then recycle the nutrients to the bioreactor.

V. REFERENCES

1. Chynoweth, D.P., "Overview. In D.P. Chynoweth and R. Isaacson (Ed.). Anaerobic Digestion of Biomass, Elsevier Applied Science, London, (1987) p. 171.

2. Rijken, B.A. Anaerobic digestion of solid organic matter. In: O. Kitanic and C.W. Hall (Ed.), Biomass Handbook. Gordon and Beach Science Publishers, New York, NY. (1989) p 319.

3. Buller, D.C., W. Parker, and J.S.G. Reid. Short-chain fatty acids as inhibitors of gibberellin-induced amylolysis in barley endosperm. Nature, 260 (1976) 169.

4. Russell, J.B. and D.B. Dombrowski. Effect of pH on efficiency of growth by pure culture of rumen bacteria in continuous culture. Appl. Environ. Microbiol. 39 (1980) 604.

5. Cheryan, M., M.A. Mehaia. Ethanol production in a membrane recycle bioreactor. Conversion of glucose using Saccharomyces cerevisiae. Process Biochem., 19, (1984) 204.

6. Abbott,B.J. and P. Gerhardt. Dialysis fermentation I. Enhanced production of salicylic acid from naphthalene by pseudonoma fluorescens. Biotechnol. Bioeng. 12 (1970) 577.

7. Carrondo, M.J.T., J.P.S.G. Crespo, and M.J. Moura. Production of propionic acid using xylose utilizing Propionibacterium. Appl. Biochem. Biotechnol., 17 (1988) 295.

8. Crespo, J.P.S.G., M.J. Moura, J.S. Almeida, and M.J.T. Carrondo. Ultrafiltration membrane cell recycle for continuous culture of Propionibacterium. J. Membrane Sci., 61 (1991) 303.

9. Xavier, A.M.R.B., L.M.D. Goncalves, J.L. Moreira,, and M.J.T. Carrondo. Operational Patterns affecting lactic acid production in ultrafiltration cell recycle bioreactor. Biotechnol.Bioeng. 45 (1995) 320.

10. Blatt, W.F., A. Dravid, A.S. Michels, and L. Nelson. Solute polarization and cake formation in membrane ultrafiltration: Causes, consequences, and control technique, in J.E. Flinn (Ed.), Membrane Sciences and Technology, Plenum, New York, NY, (1970) 47.

11. Trettin, D.R. and M.R. Doshi. In Synthetic Membranes: Hyperfiltration and ultrafiltration uses. A.F. Turbak (Ed.), American Chem. Soc., Washington, D.C., 1981.

12. Porter, M.C. Concentration polarization with membrane ultrafiltration. Ind. Eng. Chem. Prod. Res. Dev., 11, (1972) 234.

13. Bhattacharjee, C. and P.K. Bhattacharya. Flux decline analysis in ultrafiltration of kraft black liquor. J. Membrane Sci., 82, (1993) 1.

14. Treybal, R.E., Mass Transfer Operation, McGraw Hill, New York, NY. 1980)

15. Sherwood, R.K., R.L. Pigford, and C.R. Wilke, Mass Transfer, McGraw Hill, New York, NY, (1975).

16. Jönsson, A-S. Influence of shear rate on the flux during ultrafiltration of colloidal substances. J. Membrane Sci., 79, (1993) 93.

17. Althouse, J.W. and L.L. Tavlarides. Analysis of organic extractant systems for acetic acids removal for calcium magnesium acetate production. Ind. Chem. Chem. Res. 31 (1992) 1971.

18. Garcia, A.A. and C.J. King. The use of basic polymers sorbents for the recovery of acetic acid from aqueous dilute solution. Ind. Chem. Chem. Res. 28 (1989) 2104.

19. Steiber, R.W. and P. Gerhardt. Dialysis continuous process for ammonium lactate fermentation: Simulated prefermentor and cell recycling systems. Biotechnol. Bioeng. 23 (1981) 523.

20. Colon, G. and J.R. Rosenau. Lactic acid recovery from fermented acid-cheese whey via electrodialysis. Paper presented in the 1987 Annual AIChE Meeting, New York, NY.

21. Weir, A.J., B.A. Glatz, and C.E. Glatz. Recovery of propionic acid from fermentation broth by electrodialysis. Biotechnol. Prog. 8 (1992) 479.

1995 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

55-62
2745
p. 28

CRITERIA DEVELOPMENT FOR UPGRADING COMPUTER NETWORKS

Dr. Kemal Efe
Associate Professor
Center for Advanced Computer Studies
University of Southwestern Louisiana
Lafayette, Louisiana

KSC Colleague - Ray Pecaut
Communications/Networks

Contract Number NASA-NGT-60002
Supplement 19

July 28, 1995

ACKNOWLEDGMENTS

This research would not have been possible without the support of the NASA/ASEE Summer Faculty Program. **Ray Hossler** and **Karl Stiles** have made this summer program possible as well as enjoyable. **Greg Buckhingham** and other NASA personnel have made us feel at home and this tremendously helped the congenial atmosphere that was created here.

Throughout the research efforts, I benefited from intense discussions and suggestions made by **Ray Pecaut.** I am deeply grateful for his support and encouragement.

I am also grateful to **Cindy Johnson** for sharing with me significant technical information, which was no doubt the result of many year's work, for the benefit of this research.

Furthermore, I wish to extend my appreciation to **Perry Rogers** for his support and encouragement.

Finally, I am grateful to my wife and children for understanding when I had to be absent, over and above the regular work hours, for the benefit of this research.

ABSTRACT

Being an infrastructure system, the computer network has a fundamental role in the day to day activities of personnel working at KSC. It is easily appreciated that the lack of "satisfactory" network performance can have a high "cost" for KSC. Yet, this seemingly obvious concept is quite difficult to demonstrate. At what point do we say that performance is below the lowest tolerable level? How do we know when the "cost" of using the system at the current level of degraded performance exceeds the cost of upgrading it?

In this research, we consider the cost and performance factors that may have an effect in decision making in regards to upgrading computer networks. Cost factors are detailed in terms of "direct costs" and "subjective costs". Performance factors are examined in terms of "required performance" and "offered performance." Required performance is further examined by presenting a methodology for trend analysis based on applying interpolation methods to observed traffic levels. Offered performance levels are analyzed by deriving simple equations to evaluate network performance. The results are evaluated in the light of recommended upgrade policies currently in use for telephone exchange systems, similarities and differences between the two types of services are discussed.

TABLE OF CONTENTS

I. INTRODUCTION

In almost every walk of life, we are regulated by standards and limits imposed or recommended by some organization. A truck comes with a recommended load capacity. A boat or an elevator comes with its recommended number of people to be carried. Air conditioners have recommended volumes to be cooled depending on their capacity. FDA recommendations exist about how much fat, sugar, fiber, etc. one should consume per day. We can go on giving examples of such recommended limits practically forever. An area where no organization seems to have bothered with is computer networks. There are no recommendations or guidelines that advise on how much traffic to be carried on computer networks. Computer networks are now entering every organization, some serving critical functions such as hospitals and chemical plants, where network congestion may have catastrophic results. In other organizations, network congestion may not have immediate consequences, but may never the less have long term impact in operational costs or employee productivity.

The purpose of this research is to develop simple criteria that may be used in decision making in regards to upgrading computer networks. Most computer networks in use today have been installed more than 10 years ego. Due to the recent proliferation in new networking products, network managers throughout the world are faced with the difficult question of whether to continue using their old network that served them so well in the past, or whether to replace it for a new product in the market.

After much literature research and intense discussions with colleagues here at KSC, the answer to this enormous question turns out to be extremely easy and hard simultaneously, at least in the context of the KSC network. What makes it easy is a fact, that became obvious to us in a short time after starting this research, that future applications requiring network services will saturate the existing system much beyond its capacity. The hard part was due to the fact that unless the network is saturated beyond its capacity, any level of utilization below may be interpreted as being acceptable or not based purely on interpretation. There is no single indicator that dominates all other factors to favor replacing

the network for a new one. Moreover, simple management level decisions may be made to postpone the point of catastrophy, for a cost in employee productivity or the range of network services that may be supported. Ultimately, upper management must decide when it is time to upgrade the network.

II. COST/PERFORMANCE FACTORS

When making such decisions, upper management will look for a good trade-off point for cost and performance. Clearly, it is necessary to clarify the two key words: "cost" and "performance." In doing so, we take a close look at cost factors and performance factors in detail.

II.1 COST ELEMENTS:

We can divide the involved cost factors into three categories as follows:

Direct Costs: These are the direct costs with two components: (a) direct costs for maintenance and management, and (b) direct cost of personnel time (for those personnel using the network) for time lost due to slow response time.

Older technology usually has higher maintenance cost. Contrasting to the state-of-the-art technology, such costs can be predicted with reasonable accuracy. Direct cost due to slow response time can also be computed reasonably accurately since every second of employee time has a cash value based on the salary.

Subjective Costs: These are the cost factors for lost opportunity. If the existing system does not offer the necessary capacity required for state-of-the-art facilities, the inability to use such facilities may have high cost for an organization. For example, video conferencing is one area gaining significant market in recent years. Other examples include electronic marketing, or access to digital libraries, etc.

This is a relatively difficult cost category to quantify. For some organizations, the need for state-of-the-art facilities may be as real as the need for a telephone on the executive director's desk. For other companies, these may be nothing more than fancy toys. Nevertheless, there is already a sizable

customer base for these facilities. Several companies offer free access to their data bases (analogous to the 800 numbers for telephone access) so that customers can easily obtain product information, place orders, etc. In the near future, a substantial increase in the number of users is expected for remote access to such on-line data which may have been stored in graphics, video, audio, or text format. AT&T and MCI are now offering up to 1.5 Mb/s bandwidth (per connection) for 800 calls for institutions that want to offer computer access to their on-line data. Various uses of such data bases in education, commerce, etc. are anticipated.

For the purposes of this research, It is assumed that the ability of KSC personnel to access such facilities offered by others has a "high" value (even though it is difficult to quantify this value due to its subjective nature).

Cost Efficiency: The amount of bandwidth purchased per dollar spent is a measure of cost efficiency. Also, the concept of cost efficiency can be applied to maintenance, i.e., the amount of bandwidth maintained per maintenance dollar spent. As defined, the maintenance cost efficiency can be easily calculated for a network by using the maintenance records.

II.2 PERFORMANCE CONSIDERATIONS:

Performance considerations are addressed under two categories: required performance, and offered performance.

Required Performance: This refers to the level of performance needed in order to satisfy certain quality of service parameters arising from applications that are deemed necessary for an organization. The amount of required bandwidth is among the major parameters that determine the quality of service. Other parameters include the delay, the probability of packet loss, the amount of jitters tolerated, etc. To do a formal study of total network capacity needed, the relationship between these parameters needs to be well understood. For the time being, it should be interesting to consider the bandwidth requirements arising from multimedia applications, for a single user, as it will allow us to do a quick and dirty analysis for the existing FDDI backbone network at KSC. The table below lists the

requirements for video, graphics, audio, and data communications, as these are typical modes of display in multimedia. (Table is compiled from reference [1]).

	Bandwidth	Bursty vs. Continuous	Tolerable Loss	Delay/Jitters Sensitivity
CBR Video	Several* Mb/s	Continuous	10^{-6}	Tight
VBR Video	Several*Mb/s	Bursty	10^{-6}	Tight
Image	Several Mb/s	Bursty	10^{-8}	Medium
Audio	Under 100Kb/s	Continuous	10^{-2}	Tight
Text/Data	Flexible	Bursty	10^{-10}	Loose

* MPEG I specified rate is 1.5 Mbit/s for CBR video, and 2.5 Mbit/s for VBR video.

Table 1: Quality of service parameters for multimedia applications.

The specified bandwidths for CBR and VBR Video are expected to increase to 6 and 19.2 Mbit/s, respectively, in the MPEG II standard for HDTV. If we use the 1.5 Mb/s value, and since the FDDI network operates at 100 Mb/s, we can allow only about 60 users at a time engaging in a video session (e.g., to search through a manufacturer's on-line catalogue). This quick and dirty computation does not consider the other traffic normally on the network. Also it does not take into account other quality of service parameters such as delays in accessing the network.

Offered Performance: Offered performance refers to the actual quality of service offered by the existing system. Network monitoring tools exist that measure various parameters such as the number of packets sent, the length of each packet, the delay taken for a packet to reach its destination, and so on. Presumably, the management would use this information to determine the bottleneck points on the network, and make decisions about upgrading the system if and when necessary. Unfortunately there are no standards organizations that make recommendations for computer networks. However, there are voluntary standards observed by telephone companies for their switching systems. In

telephone industry, grade of service is specified as the percentage of calls lost due to busy switching system (at peak time). It appears reasonable to do a detailed study of the underlying theory behind these quality of service standards used by telephone companies and investigate their suitability for computer networks. We do so in Section 4.

III. TREND ANALYSIS FOR NETWORK UTILIZATION

In marketing literature, one curve often used is the S-curve that describes percentage of customers having purchased a given product. This curve generally looks like that in figure below. When a new product (e.g. automobile, television, computer, etc.) is introduced to the market, initially 0% of the potential customers actually have that type of product. In the first few years after the product hits the market, acceptance is low, and only a small proportion of potential buyers actually acquire that product. Eventually more and more people buy the product and the percentage of people that have bought the item begins to increase sharply. After some time, the increase in the percentage begins to level off, and reaches the 100% level gradually.

This type of curve can also be applied to model the behavior of network users, and ultimately to analyze the level of required network services. Given a network, initially we anticipate a low usage as potential users have not yet acquainted themselves with benefits and conveniences of such a service. After a while more and more people use the network causing a sharp increase in the number of users. Eventually, all people that are likely to use the network become regular users, and the proportion of people using the network levels off gradually, reaching it maximum possible level in the long run.

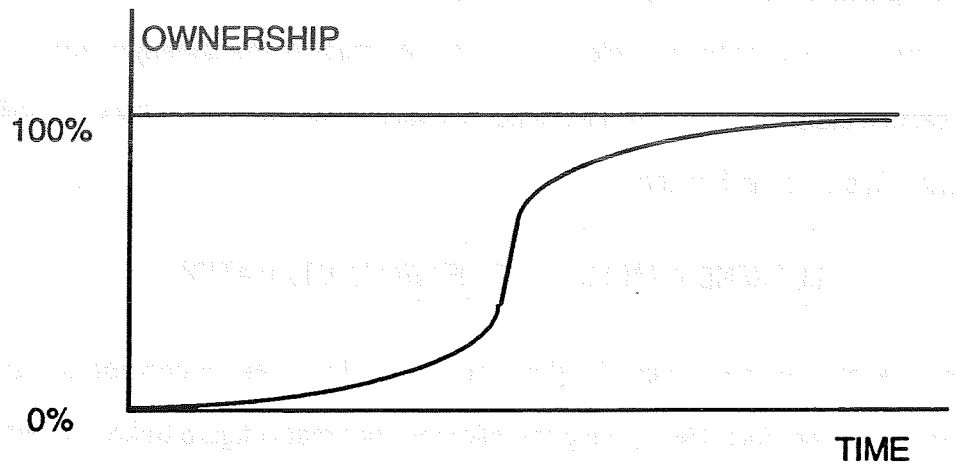


Figure 1: A typical curve representing market growth.

From the network point of view, this pattern of behavior manifests itself in the corresponding increase in network traffic. Therefore, this curve will also describe the pattern of change in network utilization and may be verified by network monitoring tools.

If we assume that this curve represents utilization patterns for computer networks, management questions to be considered are (a) what amount of traffic corresponds to the 100% level on the curve?, and (b) if that level is more than the offered capacity, then how long it will take to reach that level?.

The first question can be answered by considering the number of users, and the expected level of usage per user. Since these are difficult measures to quantify exactly, we may consider best case or worst case scenarios depending on management attitude. Best case could be defined as a conservative estimate on the usage patterns. For example, if we use the MPEG-1 bandwidth requirements for the required bandwidth per user, we can estimate the required bandwidth as 1.5N Mb/s, where N is the expected number of active users. If N=1000, then the estimated traffic would be 1.5 Gb/s. Note that N corresponds to the number of users active at any time. Thus, if for example only 20% of users are active at the same time, then 1000 active users would correspond to a total of 5000 employees. A worst case estimate can be used by considering MPEG-2 bandwidth values which is expected to be close to 20Mb/s, and assuming that a higher proportion of employees will use the network at that level. If 50% of the 5000 employees use the network at any time, requiring 20 Mb/s bandwidth each,

the total bandwidth will be 50 Gb/s. As can be seen a wide range of possibilities exist for the maximum expected traffic load, and depending on whether or not a certain application is made available to the users, different numbers may be used as the estimate for the highest traffic level.

The second question, i.e. that of how long it will take to reach a certain point on the curve, is somewhat more complicated. The reason is, although conceptually simple, this curve may come out as an aggregate function of several components, each of which having a lifetime curve by itself. The corresponding curve might look more like that in the following figure:

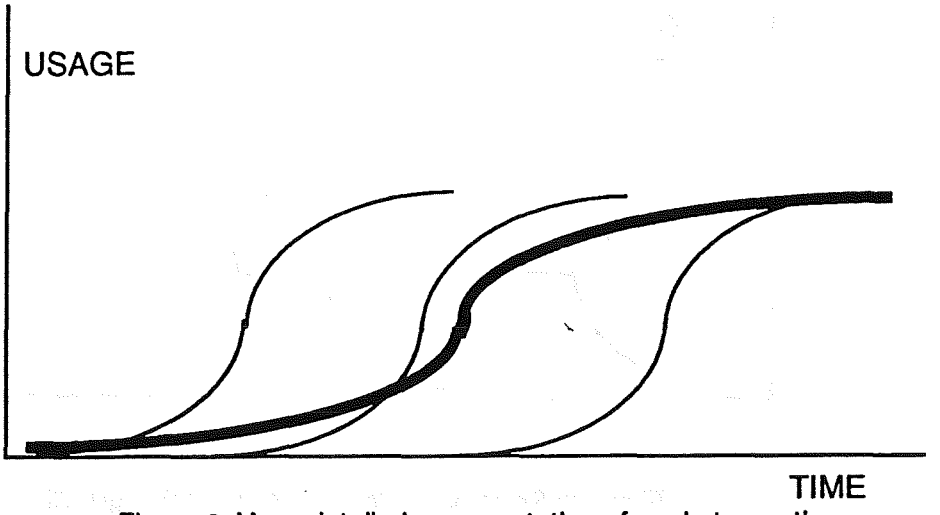


Figure 2: More detailed representation of market growth.

In this figure, the thick line represents the aggregate behavior of individual usage patterns. While at any time some of the commodities may have reached their peak usage, there may be other commodities that are in the mid-point of their usage and still others that are just entering the user domain. A detailed analysis and forecast method may therefore benefit from analyzing the usage levels of different items individually. It is very likely, for example, that regular email usage may have exceeded its mid-point while other applications such as WWW may be just beginning to claim its share of bandwidth.

III.1 FORECASTING

The question one normally asks at this point is how do we know where we are with respect to a particular curve. This is a very difficult question to answer as one can only know the past usage levels with absolute certainty. For future usage levels, we can only argue about approximate usage levels under certain assumptions. One assumption we will be making is that rates of increase do not change abruptly. In other words, we will be assuming that the lifetime curve for a commodity is not like those in the figure below.

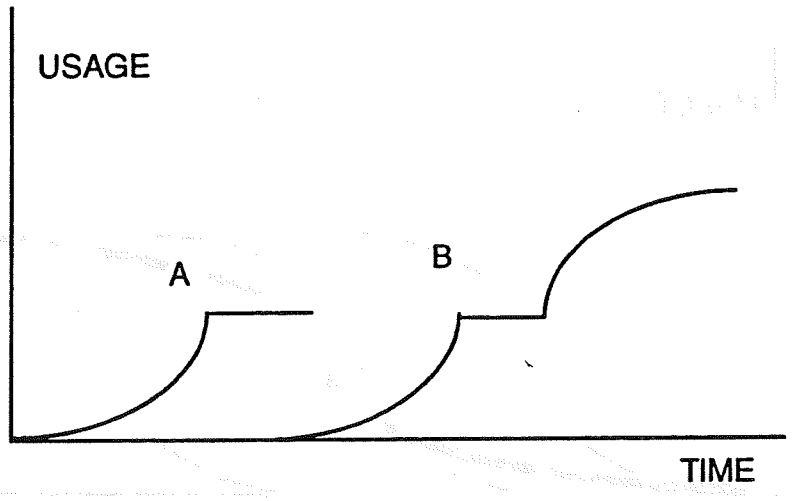


Figure 3: Some anomalous cases in market growth.

In this figure, A represents an abrupt an indefinite blockage to increase of the commodity usage and B represents a temporary blockage. Our smoothness assumption appears to be valid on intuitive grounds, but examples exist to show the contrary. For example, phenomenon analogous to that in figure A has occurred in household TV dish antennas, when cable companies objected to people having their own receiving dish antennas. The case of B occurred in cellular phone sales when a user filed law suit claiming that the cellular phone has caused brain tumor.

Having pointed out a few drawbacks of our assumption, we proceed to discussion of what can be said about the future. The assumption of forbidden abrupt change allows us to at least predict where we are on the usage curve approximately. We can argue, for instance, whether we have reached the mid-point

or not. The utility of this analysis is that it allows us to predict whether to expect significant change at usage levels or a modest change in the near future.

To explain, consider three usage levels sampled at three different times. Suppose that at times t_0, t_1, t_2, we have the measured usage levels of y_0, y_1, y_2. We can use interpolation polynomials to fit a curve that crosses the three points $(t_0, y_0), (t_1, y_1), (t_2, y_2)$. Since three points define a second degree polynomial, we can use the equation:

$$y = y_0 \frac{(t-t_1)(t-t_2)}{(t_0-t_1)(t_0-t_2)} + y_1 \frac{(t-t_0)(t-t_2)}{(t_1-t_0)(t_1-t_2)} + y_2 \frac{(t-t_0)(t-t_1)}{(t_2-t_0)(t_2-t_1)}$$

to derive the desired curve. If at present time the curve has an increasing rate, then we conclude that we have not yet reached the mid-point of the lifetime curve and thus steep increases are expected in the near future. If this curve has a decreasing rate, then the expected increase will probably begin to level-off, and future increases will not be as steep as what has been observed in the past.

Example: Traffic readings on the KSC backbone network has shown that the utilization has doubled every year for the last two years. If we denote the utilization of two years ego as u, then last year's utilization is 2u, and the current utilization is 4u. If time of two years ego is $t_0 = 0$, then last year is $t_1 = 1$, and this year is $t_2 = 2$. Using these, we evaluate the above equation at points

$$(t_0, y_0) = (0, u);$$
$$(t_1, y_1) = (1, 2u);$$
$$(t_2, y_2) = (2, 4u);$$

and we obtain:

$$y = u(\frac{t^2}{2} + \frac{t}{2} + 1).$$

This equation has the derivative:

$$y' = u\left(t + \frac{1}{2}\right)$$

which is positive for every positive value of u and t. Evaluated at time t=2, the slope is y'=(5/2)u, where for t=1, it is y'=(3/2)u. This means that the increase now has a higher rate than one year ego, and even if this rate does not change in future, we can expect steep increases in the volume of communication on the network. Moreover, we also conclude that the current utilization rate is below the 50% level of the final saturation level of user demand, as beyond the 50% level the rate of increase would decline instead of increase.

IV. NETWORK PERFORMANCE ANALYSIS

The above model of trend analysis focuses on the required network usage by the user population, and does not say anything about whether or not the available system will be able to provide the required bandwidth to satisfy user community. To predict if the network will still satisfy the user needs in future, we need a method of performance analysis for the available system.

The system of computer networks at KSC is probably one of the most complex networks in the United States. It serves more than 10,000 users under three administrative authorities, using nearly 150 local network segments. All of the major vendors and networking protocols are represented, including Appletalk, Novel, Ethernet, token ring, DECnet, etc. These networks are connected together by three major backbone local area networks, which are in turn connected through the Kennedy metropolitan area network. A second metropolitan area network is dedicated to connect these three major units to other NASA centers.

No method of analysis, whether analytical or simulation-based, can handle the system of networks at KSC as a whole. While analytical models in the literature can yield very accurate results when a sufficiently detailed model of the network is considered, these models are mostly specialized for a single network running under a single protocol. Simulation models generally allow even more accurate representation of the real world environments, but they are inefficient and take a long time to run even for simulating a single network protocol in isolation. When nearly one hundred networks connected through multiple backbones is considered, the required running time of the simulation program will be prohibitive.Here we present a simple method that can be used for estimating

performance bounds in any network regardless of the communication protocol. The major advantage of the method is its remarkable simplicity.

Consider some network with 1 Mb/s bandwidth, 1000 users connected to it, and average frame size 1000 bits. Clearly, if each user generates data at rate exceeding 1 frame per second, then the total offered load exceeds the network bandwidth of 1 Mb/s, and the delay at each station will grow without bound. Below this level of traffic, we identify two regions of delay: a region of low delay where capacity is more than adequate to handle the offered load, and a region of high delay where the network becomes a bottleneck (i.e. a source of noticeable delay). The crucial question is: What region will the network operate in, based on projected offered load. To develop a method of analysis, we define five quantities:

X : the average time required to transmit one frame, once medium access has been gained by a user.

t : mean time that a user spends between two frames generated by that user.

D : system delay, which is the sum of queuing and transmission delays experienced by an average frame.

S : mean total throughput (number of frames sent per unit time) on the network for all users.

N : the number of users having direct access to the network.

Here we wish to use t and X as the input variables and compute upper bounds on S and N, and a lower bound on D. Before proceeding further, we also define two intermediate values in terms of the input variables t and X.

Define:

$$\lambda = \frac{1}{t}$$

$$\rho = \lambda X$$

Intuitively, λ measures the arrival rate of frames at each user cite, since a user generates one frame every t seconds. Also, ρ measures the utilization by a user of the network. It equals to the arrival rate, normalized by the amount of time taken to send one frame.

Having defined some system parameters, we can readily make some general comments about recommended usage levels for computer networks. To do so, consider the general behavior of a queuing system as shown below.

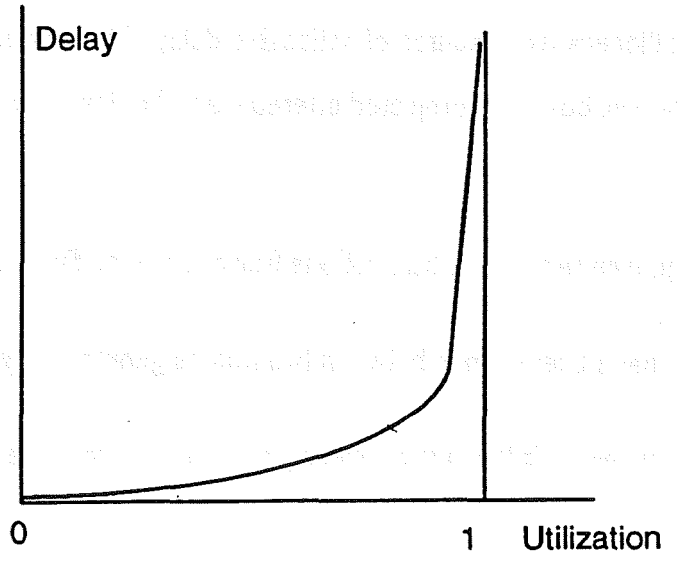


Figure 4: Typical delay curve for computer networks.

As shown, delay tends to infinity as utilization tends to 100% level. Below this level, delay may appear small, but such values may be deceiving. For example, we will show later in the case of an FDDI network that for about 50% utilization the mean delay could be around 40 microseconds. Yet the utilization level giving this average behavior frequently reaches 100% level giving rise to frequent occurrences of network congestion. The reason for this abnormal behavior can be easily explained by the Poisson property. Arrivals to a queue following poisson property are extremely irregular, so much that if the mean value of the inter-arrival times is t, its variance is t^2. (A good source is [2]).

It follows from this observation that, if t is at a level that will yield 50% utilization on the network, there are frequent occurrences of extremely long queues. To avoid frequent congestion that would arise from this phenomenon, it is necessary to keep the utilization level below 50%.

IV.1 Approximate methods for multi-access (shared) networks:

Here we give approximation methods for throughput and delay analysis for any network. In the appendix we will give extremely simple methods to derive the exact values for delay for some specific cases. To find an upper bound on the total throughput, assume that t > 1, and thus all the messages generated by a user are transmitted with some bounded delay. Then the throughput for one user is just S = λ = 1/t. Since each of the N users have this amount of traffic, the total throughput is:

$$S = \frac{N}{t}.$$

This upper bound cannot increase more than the total network capacity, which is:

$$S \leq \frac{1}{X}.$$

Thus, equating the right had sides of the two equations above and solving for N, we find that:

$$N \leq \frac{t}{X}$$

or, using t = 1/ λ, we have:

$$N \leq \frac{1}{\lambda X} = \frac{1}{\rho}.$$

When the equality is satisfied the network is congested at the saturation point. For N less than this value the network is able to transmit all the frames submitted to it, but the amount of delay may be large if N is close to its maximum value.

We can compute a lower bound on delay by using M/G/1 approximation. The justification for the approximation is that the users can at best have access to the network like there is no interference due to network access protocols. Let

$$\rho = \frac{NX}{t}$$

denote the utilization by all users of the network. In terms of ρ delay for M/G/1 queue is given as:

$$D = X\left(1 + \frac{\rho}{1-\rho}\right);$$

for exponentially distributed service times, and,

$$D = X\left(1 + \frac{\rho}{2(1-\rho)}\right);$$

for deterministic service times (i.e. all customers have equal service time). The first of these two formulae may be used when the frame size is variable for a network, and the second may be used when the packet sizes are fixed.

An interesting property of these delay equations is that they are both in the form

$$D = Xf(\rho).$$

In other words, both of these are equal to X multiplied by some function of utilization. Since X is the time that would be needed to send an average frame, once a user gains control of the network, the function $f(\rho)$ may be used directly to obtain a comparative measure of the delay.

Here we give some examples to illustrate the use of these equations.

Example 1: A workstation is attached to a 1-Mb/s network, and it generates, on average, three frames per minute, with messages averaging 500 bits. Thus, message transmission time is $X = 500\mu s$, and the mean inter-arrival time is t=20s. The maximum number of workstations that can be supported is:

$$N = \frac{t}{X} = \frac{20}{500 \times 10^{-6}} = 40{,}000.$$

Example 2: On a 100-Mb/s network, each user generates 30 frames of 36000 bits every second. This situation would arise, for instance, if video information is sent over the network. Thus $X = 36 \times 10^{-5}$ seconds, with $t = 33.33 \times 10^{-3}$ seconds. Then, using the above formula, we find that N=92.

This example illustrates that at most 92 users can use the network at the same time. In an organization, it may be reasonable to assume that only about 5-10% of users are engaged in video communication at any time. Using the 5% level, we find that total number of employees in such an organization should not exceed $92 \times 20 = 1840$.

Example 3: For the above example, suppose there are 90 users actively engaged in video communication. Then,

$$\rho = \frac{90 \times 36 \times 10^{-5}}{33.33 \times 10^{-3}} = 0.972,$$

and, by using the second delay formula, the corresponding maximum delay will be at least:

$$D = X \times \left(1 + \frac{0.972}{2(1-0.972)}\right) = X \times 18.41.$$

By using the above value for X, we can easily compute that delay is at least D=6.63 milliseconds. While this delay may appear insignificant at first, we remind the reader that it represents at least 18 times the delay that would occur on a lightly utilized network. This indicates that the network is very close to its total congestion state.

For the purpose of approximate calculations X may be replaced for a larger value X' if we wish to account for overheads due to network protocols. The following example illustrates this point.

Example 4: In an FDDI network, the end of each packet is indicated by a free token. Assume that the number of bits in the token is negligible compared to the total frame length. However, two consecutive stations are 10 Km apart, which means that time taken for the next workstation to know that it can send messages is equal to the token flight time. If we assume that network propagation delay is 1' per nanosecond, then we have $X_{ov} = 3.28 \times 10^{-5}$ seconds. Since $X' = X + X_{ov}$, replacing every occurrence of X for X', and performing the computation as above, we find that only 84 users will actually saturate the network. Here we assumed that each time the token is obtained, a user sends just one packet, even if there are many waiting to be sent.

A more detailed treatment of delay for FDDI is given in the appendix.

IV.2 Methods for switched systems:

In telephone circuits, the criterion used to upgrade the system is as follows [3]. Traffic levels are monitored continuously, and peak hours of utilization are determined. Traffic for 10 of the busiest hours (30 hours in Europe) occurring on distinct days are averaged and the corresponding probability that an arriving customer finding all the switches busy is computed. If this probability is between 0.001 and 0.01, then grade of service is considered good. If the probability exceeds 0.01, then more switches are added to the system.

To adopt a similar approach for computer networks it is necessary to have a model that is applicable to both multi-access systems and to switched systems. For this purpose, we can model the FDDI network as a time division multiplexer. If each user required 1Mb/s bandwidth, we divide the total bandwidth of 100 Mb/s into 100 switches. If each user requires 20 Mb/s, this corresponds to having 5 switches. With this analogy, we can compute the utilization level needed for the switching system to give 1% probability of queuing. Having found the utilization level, we can then compute the corresponding delay for the FDDI network. This should give a good idea about the kind of delay faced if we used the same criterion as those of telephone systems.

In queuing terms, switched systems can be modeled as M/M/m queuing systems and solved by the help of Markov Chains. Unfortunately, there appears to be no simple way to explain these models, thus in this section we just summarize the main results and explain how these formulas may be used.

The M/M/m system is a queue with m servers. Each server may correspond to a switch in a telephone exchange system. Alternatively, for data networks such as ATM, each server corresponds to an input port.

Define: $\rho = \frac{N\lambda X}{m}$

Here N is the number of users, λ is the arrival rate of messages generated by each user, X is the service time for each call (or packet), and m is the number of switches. For an arriving customer (or packet, or telephone call) probability of finding all servers busy, thus having to wait is:

$$P_Q = \frac{1}{1+(1-\rho)\frac{\sum_{k=0}^{m-1}\frac{(m\rho)^k}{k!}}{\frac{(m\rho)^m}{m!}}}.$$

The number of frames waiting is: $N_Q = P_Q \frac{\rho}{1-\rho}$,

and the corresponding waiting time is:

$$W = \frac{N_Q}{\lambda} = \frac{N_Q}{N} = \frac{P_Q \rho}{N(1-\rho)}.$$

These are well known equations originally due to Erlang. For a given value of P_Q it is not possible to solve for the corresponding value of ρ since P_Q is a complex function of m and ρ. We therefore write computer programs which solve for the desired values by iterative methods. Sample results of these algorithms are given below.

# Of switches	Utilization (pr.=0.01)	Utilization (pr.=0.05)	Utilization (pr.=0.1)
10	0.409	0.529	0.599
20	0.550	0.651	0.706
30	0.621	0.709	0.757
40	0.665	0.745	0.788
50	0.697	0.770	0.809
60	0.721	0.789	0.825

Table 2: Utilization values for various grade-of-service parameters and different system sizes.

As can be observed from this table, the utilization values increase quite high for increasing number of switches regardless of what P_Q value is used. This shows that, by the time telephone industry reaches P_Q =0.01, they are indeed receiving quite high utilization levels. This is particularly true for very large number of switches, (which is the case normally in industrial settings) when utilization approaches close to 100% even for P_Q =0.01. We have already pointed out that utilization levels above 50% are not advisable for multi-access data networks. This shows that the criteria used in telephone industry is not applicable for multi-access networks.

As an example of this, we illustrate a hypothetical situation of using the FDDI network for multimedia applications. First, recall that maximum packet length for FDDI is 36000 bits. Assume that part of the computer screen of size 512 by 512 is used for video, and the rest is used for text. The video part, in 8-bit color and 32 to 1 compression ratio comes to about 30,000 bits. Thus each screen nicely fits in a

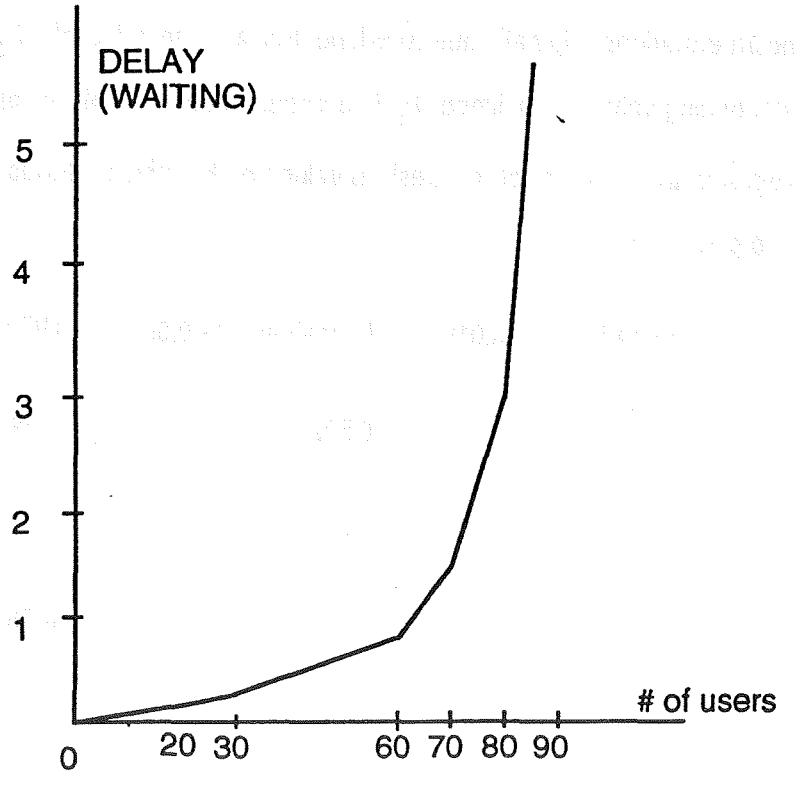


Figure 5: Delay v.s. utilization in FDDI (M/G/1 approximation with deterministic service times).

single FDDI packet, leaving ample room for the accompanying text. As a result, we assume that each user sends 30 such packets every second (corresponding to refresh rates for video pictures). The question we want to answer is how many users can be supported.

Using the delay equation for M/G/1 case (which is a generous model for FDDI network) we plotted the delay as a function of increasing number of users as shown above.

Here we used the deterministic equation since all the packets have equal length of 36000 bits. The vertical axis shows the delay in multiples of X, which is, as before, the time taken to send a single packet. X itself is very small, in the range of a few tens of microseconds. Never the less, as the number of users reaches around 90, the multiplication factor on the vertical axis reaches close to infinity. Thus, using the 50% rule of thumb, we should not really try to support more than about 45 users on that network. Contrasting the situation with the switched systems, we see that for 60 switches we reach the probability $P_Q = 0.01$ at about 72% utilization. This is clearly a forbidden region to operate the FDDI network.

V. CONCLUSIONS AND RECOMMENDATIONS

This research has addressed the problem of developing practical criteria to upgrade computer networks. The problem turns out to be extremely easy and difficult simultaneously. It is easy when network traffic generated by those applications deemed desirable far exceeds the network capacity. It is difficult, however, when the current utilization levels are far below the network saturation point.. For this reason, we suggested various cost/performance factors that may be used to asses the desrability of increasing the offered bandwidth. We also suggested methods for forecasting future traffic patterns and interpolation methods to determine the current position in a demand curve. Simplified analytical tools are described and examplified to predict network behavior under expected future utilization levels. Some specific recommendations resulting from this research include:

1) Network utilization at or above 50% levels is not recommended due to frequent congestion.

2) A method of monitoring network traffic based on application is desirable. The obtained data can better represent expected traffic growth by different applications, leading to more accurate predictions.

3) A traffic measurement method similat to that used in telephone industry is advisable. Average traffic levels can be very low, yet the traffic at busy time is a more accurate representation of what the users experience. It is advisable to record the 10 busiest traffic levels in a year for use in decision making.

4) It is advisable to use the forecasting method presented in this paper, based on past data, to see how accurately it reflects the future, using the "future" data collected later.

VI. REFERENCES:

1. J. Y. Hui, J. Zhang, and J. Li; "Quality-of-Service Control in GRAMS for ATM Local Area Networks," *IEEE Jour. SAC,* 4 (13), 700-709.

2. D. Bertsekas and R. Gallager, Data Networks, (second edition), Prentice hall, 1992.

3. G. Held, Data Communications Networking Devices, (third edition), John Wiley & Sons, 1992.

APPENDIX

Derivation of M/G/1 delay equation:

Here we offer a simple minded but exact derivation for delay equations considered in this paper, in order to provide further understanding of how these equations may be used. Consider the different components of delay in a queue with a single server. Specifically, we focus on the time W spent waiting in the queue. Once we compute W, the total delay is just:

$$D = X + W.$$

The waiting time consists of:

$$W = R + XN_Q$$

where R is the remaining time for the frame being serviced, X is the average time needed to send one frame, and N_Q is the number of frames waiting to be serviced ahead of the arriving frame. From Little's law, we know that N_Q is equal to throughput times delay, that is, $N_Q = \lambda W$. Using this in the above equation and solving for W, we find that

$$W = \frac{R}{1-\rho}.$$

Now it remains to compute R, the expected value of the remaining service time for the frame being transmitted. If all the messages have equal length (deterministic service time assumption), and the transmission has just started, then the remaining time is obviously X. If transmission is about to finish, then the remaining time is zero. The average value is thus X/2, but we must also consider the fact that this delay is experienced only while the served is busy (an even which occurs with probability ρ). If the server is not busy (which happens with probability 1- ρ) the term R is equal to zero. Thus we have

$$R = \rho\frac{X}{2} + (1-\rho)\times 0$$
$$= \frac{\rho X}{2}.$$

Using this in the above formula, for systems in which all frames have equal length, we find that

$$W = \frac{X\rho}{2(1-\rho)}.$$

If the frame lengths are not equal, then we need to know the distribution of frame lengths in order to derive the value of R above. From practical observations, it has been found that an exponential distribution can be used to represent the distribution of frame lengths in data networks. One interesting property of exponential distribution is its "memoryless" property. This means that no matter how long time has been already spent servicing a customer (or transmitting a frame in a data network) the expected value of the remaining time is statistically equal to the mean service time (or mean frame length). Many examples of such phenomena exist in real life. The most famous example is the light bulb; no matter how long a light bulb has been in use, the expected remaining time it will last before burning out is equal to the mean lifetime over all light bulbs.

With these consideration, we expect that, given that the system is busy, the length of time needed to finish a frame in transmission is just X instead of X/2. in other words, we have $R = \rho X$, and using this we find

$$W = \frac{X\rho}{1-\rho}.$$

Delay equation for FDDI network:

So far, we assumed that the amount of overhead due to network protocols is negligible. If this overhead is known, or can be approximated, then we can also replace X for X', where $X' = X + X_{ov}$. This reasoning assumes that every packet sent experiences the overhead, and such a method can be applied for and FDDI network where each time a station acquires the token it sends only one packet. In an exhaustive algorithm where a station empties its queue, the situation is more complicated. Below we treat this case in detail to give an idea about how to approach such problems.

Consider the equations we have seen before for the M/G/1 queue:

$$D = X + W$$
$$W = R + X \times N_Q.$$

In FDDI Token Ring with exhaustive policy an arriving frame waits for the time periods of R units for the current frame to finish transmission plus an additional time needed to acquire the token. That is,

$$W = R + V + X \times N_Q.$$

where V is the time needed to acquire the token, and the rest of the terms are as before. Here we do not alter the term $X \times N_Q$ since once the token has been received all the waiting frames are transmitted. We just need to add the term V to R, since R is now larger than what it would have been if the user had not waited for the token.

To compute V, let v denote the time needed to pass the token from one user to the next. In an FDDI network, this time can be approximated by the network propagation time since the idle token immediately follows the last message sent by a user. Suppose that there are N users with equal distance from one another. Then, total propagation time of the token around the network is Nv. In the worst case, a new frame arrives just after the local user has finished passing the token to its neighbor, and the new frame may need to wait for N-1 other users before it obtains the token again. In the best case, token may be already at the local station, thus no time is needed for token rotation. The average case is, however, the case of waiting for (N-1)/2 users, but all of this wait occurs only if the network is utilized by some user. This happens with probability ρ. If the network is not being used, an event which occurs with probability 1- ρ, then the arriving frame may arrive just as the token arrives to that station but has no chance of being transmitted, and its average waiting time will be Nv/2. Expressing these quantities in more formal terms, we have

$$V = \rho \frac{(N-1)v}{2} + (1-\rho)\frac{Nv}{2}.$$

After simplification, we obtain

$$V = \frac{(N-\rho)v}{2}.$$

Since we already knew how to compute the rest of the terms in delay equation (e.g. the terms R and $X \times N_Q$), combining all of these terms we obtain

$$W = \frac{X\rho}{1-\rho} + \frac{(N-\rho)v}{2(1-\rho)}.$$

This is the waiting time expression for an FDDI network when packet lengths are distributed exponentially. We simply divide the first term by 2 when packet lengths are equal. Justification for this is similar to that we gave for the M/G/1 queue. Note that, throughout the appendix we defined $\rho = \lambda X$.

1995 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

APPLICATIONS OF DECISION ANALYSIS AND RELATED TECHNIQUES TO INDUSTRIAL ENGINEERING PROBLEMS AT KSC

56-81
7746
P-46

Dr. Gerald W. Evans
Professor
Industrial Engineering Department
University of Louisville
Louisville, Kentucky

KSC Colleague - Tim Barth
Industrial Engineering

Contract Number NASA-NGT-60002
Supplement 19

August 4, 1995

ACKNOWLEDGMENTS

I'd like to thank Ms. Kari Stiles and Professor Ray Hosler of the University of Centraal Florida for their administration of the 1995 NASA/ASEE Summer Faculty Fellowship Program. Their attention to detail and the time they spent in arranging various tours and social programs are much appreciated.

The staff of the Industrial Engineering Group at KSC: Tim Barth, JoAnn Leotta, and Amanda Mitskevich provided valuable direction and advice for my work this summer. They are a dedicated group whose work should prove to be very valuable to KSC in the years ahead.

In addition, I'd like to thank Eric Redding for his hospitality, and Kevin Mellet, Cedric Hill, Shelby Roberts, and the other thermal processing system engineers for their help in my understanding of the tile manufacturing process.

Finally, I'd like to thank Linda Tarrence of the University of Louisville and Terrie Bumch of the Kennedy Space Center for typing this report.

ABSTRACT

This report provides:

1. A discussion of the origination of decision analysis problems (well-structured problems) from ill-structured problems.

2. A review of the various methodologies and software packages for decision analysis and related problem areas.

3. A discussion of how the characteristics of a decision analysis problem affect the choice of modeling methodologies, thus providing a guide as to when to choose a particular methodology.

4. Examples of applications of decision analysis to particular problems encountered by the IE Group at KSC.

With respect to the specific applications at KSC, particular emphasis is placed on the use of the Demos software package (Lumina Decision Systems, 1993).

SUMMARY

A problem can be defined as a situation in which there is a gap between what is and what should be. Most problems in business and industry originate as ill-structured problems--that is, problems in which the criteria, stakeholders, alternative solutions, etc. are not immediately obvious. A problem formulation process (e.g., involving a procedure such as the Why-What's Stopping Technique) is helpful in defining well-structured problems, in which the criteria, stakeholders, etc., are clearly defined. These well-structured problems originate from ill-structured problems.

A well-structured problem is a decision problem in which one has several alternative problem solutions, each leading to a different outcome (as defined by the performance measures of the situation). The problem then is to select and implement the best alternative solution.

The field of operations research/management science (OR/MS) provides a variety of models and techniques for solving well-structured problems: simulation models, analytic models, gaming models, payoff tables, judgmental models, influence diagrams, decision trees, mathematical programming, goal programming, multiattribute value functions, multiattribute utility functions, and the analytic hierarchy process, to name a few. By listing the characteristics of a problem, relating to the following descriptors: problem importance, number and types of decision makers, number of performance measures, complexity of functional relationships between alternative solutions and the performance measure space, and the amounts of uncertainty/risk associated with various alternative solutions; one can develop guidelines for the selection of appropriate OR/MS techniques.

Many of these OR/MS techniques can be useful tools for the Industrial Engineering Group at the Kennedy Space Center. In particular, those techniques which allow for the consideration of uncertainty/risk in the decision making process (often termed decision analysis techniques) are appropriate for a variety of problem situations. Example applications of these techniques presented in this report include the evaluation and ranking of Phase I SBIR proposals, OMS heater modification benefits analysis, assessment and ranking of proposed modifications of orbiter processing procedures, and the splash vs. NC process decision for the manufacture of orbiter tiles. Implementation of appropriate mathematical models for those decision situations is accomplished through the use of the Demos software package.

TABLE OF CONTENTS

LIST OF FIGURES

1. INTRODUCTION

A problem can be defined as "any situation in which a gap is perceived to exist between what is and what should be," (VanGundy, 1988). One way that a problem can be classified is by the amount of structure contained within it. That is, a problem can be well-structured, semi-structured, or ill-structured. A well-structured problem in characterized by the fact that all of the information required to generate a good solution to the problem is readily available. Such information includes explicit, well-defined criteria, well-defined alternative solutions, etc. An example of a well-structured problem might be the following:

> Management has asked you to determine whether an orbiter processing modification should be implemented. The simple criterion specified is a required payback period of one year.

Note that in this problem, the alternatives (implement the modification, do not implement the modification) and the evaluation criterion (required payback period of one year) are well-defined.

Most problems encountered in business and industry are ill-structured problems. That is, the criteria involved in selecting a solution are not explicit, and the various alternative solutions are not easy to come by. An example of an ill-structured problem might involve something like the following situation.

> Management has determined that there have been too many incidents involving safety. They have asked you to look into the problem.

Note that in this situation the alternatives and the criteria to use for evaluating alternatives are not immediately obvious. Hence, this is an ill-structured problem.

Operations Research/Management Science is basically concerned with solving well-structured problems. A variety of methodologies, procedures, techniques, algorithms, software packages, etc. have been developed to solve such problems.

The purposes of this report are to illustrate how well-structured problems can be formulated from ill-structured problems; to provide a taxonomy by which problems can be classified; to relate this taxonomy to the variety of OR/MS modeling methodologies and software packages available for use; and to provide example applications for the Industrial Engineering group at the Kennedy Space Center (KSC). The relationship provided between problem taxonomy and OR/MS methodologies should provide the IE group with a set of rough guidelines for the selection of modeling methodologies for solving problems. The emphasis will be on decision analysis methodologies, which explicitly consider the uncertainty/risk aspect of problem solving.

The next section of this report briefly describes the elements associated with problem and their associated solution methodologies. The third section of this report discusses problem formulation/definition, as a way of constructing well-structured problems from ill-structured

problems. Particular emphasis is placed on the Why-What's Stopping Technique, and on the process of structuring objectives and attributes. The fourth section of the report briefly describes various modeling methodologies and techniques, and associated software packages for solving well-structured problems. The fifth section lists and discusses the important characteristics of well-structured problems, especially as they relate to the selection of appropriate modeling methodologies, software packages, etc., listed in the fourth section of the report.

Examples of the uses of some of the techniques for solving various decision problems at the Kennedy Space Center is given in the sixth section of the report. Particular emphasis is placed on applications involving the use of a relatively new software package, Demos, recently acquired by the Industrial Engineering Department. Finally, the last section of the report provides a summary.

2. ELEMENTS OF A PROBLEM AND AN ASSOCIATED SOLUTION METHODOLOGY

The elements of a problem and a problem solution methodology can be defined according to the following groups of entities:

1. Decision Makers, Stakeholders, and Analysts.
2. Performance Measures, Objectives, Goals, and Attributes.
3. Alternative (Solutions), Decision Variables, and Design Variables.
4. Constraints.
5. Future States of Nature.
6. Model(s) for Evaluation.
7. Ranking/Optimization Models and Techniques.
8. Implementation Issues.

A decision maker is the person responsible for making the decision about the best course of action (i.e., alternative) for solving a problem. Typically, a problem will have several decision makers. As an example, a group of corporate executives may be decision makers with respect to determining on which projects a comparing should bid.

A stakeholder is one who will be affected by the problem and its ultimate solution, but has no direct say in how the problem should be solved. As such, decision makers should take into account the goals and objectives of the stakeholders in solving a problem. An example of a stakeholder would be someone who owns shares of stock in the company referred to above.

An analyst is a technical expert in solving problems. As such, an analyst helps the decision maker in the structuring of performance measures, and the development of models for problem solution.

Performance measures, objectives, goals, and attributes are all things that can be used to rank alternatives for problem solution. As examples, an objective represents a "direction" for improvement, e.g., minimize cost. A goal represents a desirable level with respect to an objective: obtain a cost of less than $10,000. An attribute represents a measure for an objective: yearly cost is thousands of dollars.

Typically, a problem has several conflicting objectives which must be considered in the solution process. For example, objectives relating to quality typically conflict with objectives relating to cost. The reasons for these conflicts have to do with the multidimensional nature of products, processes, etc., as well as the varying types of stakeholders which may be associated with a problem.

Attributes may be quantitative or qualitative in nature. For example, prestige on a scale of 0 to 10, with "10" meaning "world-renowned prestige," would be qualitative in nature; yearly cost in thousands of dollars would be a quantitative attribute. In addition, objectives may also be measured by proxy attributes. An example of this would be measuring quality of ambulance

service through the use of response time in minutes. Proxy attributes are often used when an objective is difficult to measure exactly; hence, a proxy attribute measures the degree of achievement of an objective in an indirect fashion.

Alternatives, or alternative solutions, are just different specific approaches for solving a problem. For example, one approach for solving a problem with a factory's production rate would be to implement a new inventory system. Of course, this solution could lead to another problem of choosing the parameter values for the inventory system.

Sometimes complete alternatives are specified as combinations of alternatives. For example, a company may have up to 10 different projects that it can undertake. It can chose none, one, two, etc., or all 10 projects. The combination of projects to undertake is the alternative, and since there are a large number of combinations, there will be a large number of alternatives.

In addition to situations where one would have just a few (e.g., two to ten) alternatives and a large number of alternatives, (e.g., hundreds, thousands, or millions), there are problem situations involving an infinite number of alternatives. This can occur, for example, when one has continuous decision variables. An example of this would be when one must set the velocity of a conveyor.

Constraints are used to restrict the alternatives that one can consider for a problem situation. For example, alternative solutions to a problem may be restricted by the fact that no alternative may have an initial cost of greater than five million dollars, regardless of the future profits resulting from that alternative.

Future states of nature can basically be represented as parameter values, over which the decision maker has no control, that can affect the outcome (i.e., performance measure values) associated with an alternative. For example, the cost savings associated with a suggested change in a maintenance activity for the orbiter is dependent upon the number of times that the activity is to take place, which, in turn, is dependent upon the number of orbiter flows. Since the number of orbiter flows is <u>not</u> under the control of the decision maker, it is an uncertain quantity, and its value does affect the outcome (i.e., cost savings) associated with implementing the change, it is a state of nature.

One of the primary tasks associated with problem solving is identifying the future states of nature and determining the likelihoods (i.e., probabilities) associated with these states. The uncertainty associated with the state of nature can result in uncertainty associated with the outcome associated with an alternative.

A model for evaluation is basically a way to map alternative solutions into the outcome space (i.e., attribute values). The outcome associate with the alternative may be either deterministic or probabilistic in nature, depending on the states of nature that exist. A wide variety of modeling methodologies are available for evaluation, including analytic models, simulation models, gaming models, and judgmental models (Miser and Quade, 1985). An example of part

of an analytic model would be an equation which relates the cost associated with performing some procedure to the number of times that the procedure is performed.

The evaluation model used for a particular situation is dependent upon the complexity of the relationships between the alternative chosen and the attribute values. For example, in a factory, the relationship between the numbers of machines in various work centers and the production rate is a complex one. In order to establish this type of relationship, one must usually develop a simulation model of the factory under study.

As another example, the relationship between the number of direct man-hours saved from the elimination of some procedure and the total cost savings associated with the elimination could probably be written in the form of a single linear equation:

$$S = atx + bmtx$$

where

S = total cost savings
x = number of hours saved from the elimination of the procedure
t = the number of times that the procedures would have been performed
a = the cost per hour for direct manhours
m = the number of indirect manhours required per direct manhours
b = the cost per hour for indirect manhours

Ranking/optimization models and techniques are used to rank alternative outcomes and /or choose the best feasible outcome (associated with a particular alternative). These models/techniques are sometimes "part of" evaluation models. For example, a linear program is a model which evaluates; and, with the revised simplex method as an optimization technique, allows one to find an optimal solution to the linear program. "Sophisticated" ranking/optimization models and techniques are only needed when either of the following two condition exist:

1. There are too many alternatives involved (e.g., as a result of the use of continuous decision variables or combinations) to allow a complete evaluation of each of them.
2. The evaluation model allows for multi-attributed output and/or uncertainty in the outcome.

When neither of the above two conditions holds, then the ranking of alternative outcomes is a simple process. For example, if the sole criterion or attribute is to maximize the predicted (deterministic) net present worth of the chosen alternative, and there are only four alternatives to consider, then the alternatives can just be ranked in order of decreasing net present worth.

If both conditions listed above exist in a problem situation, then two types of processes are

required of ranking/optimization models and techniques: an optimization process (for the first condition) and a preference structuring process (for the second condition). The optimization function is required when there are just too many alternatives to allow a complete evaluation of each alternative. Optimization techniques are typically associated with mathematical programs--linear programs, nonlinear programs, integer programs, etc. A preference structuring function is required when one has multiple, conflicting objectives to consider and/or probabilistic outcomes. Examples of models/techniques which one can employ in these areas include the scorecard approach; payoff tables; various optimization techniques such as revised simplex, interior point methods, branch-and-bound, dynamic programming, etc.; lexicographic ordering; goal programming; multiobjective mathematical programming; multiattribute value functions; multiattribute utility functions; and, the analytic hierarchy process. Ranking/optimization models and techniques are discussed in greater detail in Section IV of this report.

Finally, issues involving implementation of a selected alternative are often neglected. In fact, little attention is given to this important area is the OR/MS literature. Suffice it to say, however, that the implementation process becomes much simpler when one has a good solution (as a result of a good problem formulation) and has kept the decision makers heavily involved through the problem formulation and solution processes.

3. FORMULATING WELL-STRUCTURED PROBLEMS FROM ILL-STRUCTURED PROBLEMS

The steps associated with solving an ill-structured problem are (page 122 of Miser and Quade, 1985):

1. Formulation/definition of the problem.
2. Generation of alternatives.
3. Forecasting future "states of nature".
4. Identifying the consequences associated with the alternatives.
5. Comparing and ranking alternatives.
6. Implementation of the selected alternative(s).

The step which results in most of the difficulty associated with problem solution is problem formulation/definition. That is, most people in solving ill-structured problems do not correctly identify the problem, and hence, their "solution" does not address the real problem. Instead, what's often done is that a problem symptom is identified as a problem. Hence, correct problem formulation is a crucial, perhaps the most crucial, aspect of problem solving. As noted by Ackoff (page 8, 1974):

> "We fail more often because we solve the wrong problem than because we get the wrong solution to the right problem."

Formulation of a problem will provide, among other things (Miser and Quade, 1985):

1. A list of decision makers and stakeholders.
2. A preliminary list of objectives and attributes.
3. A specification of some promising alternatives.
4. A definition of at least some of the constraints.
5. A specification of the states of nature.
6. Specification of the types of evaluation and criterion models that one might use.
7. A plan for the analysis.

Note that some of these specifications may be very general in nature. For example, one alternative for solving a problem in a manufacturing system might be to install a new inventory control system. The specification of the parameters of the system might not come until later, or not at all, depending on whether or not the alternative was later rejected from consideration. The point here is that the problem formulation stage of the problem solution process is dynamic in nature. Objectives/attributes can be revised, alternatives can be more specifically defined, constraints can be relaxed, etc.

The problem formulation stage should be approached cautiously, because the aim is to develop an appreciation for the problem context, and to not impose a rigid structure on the problem (page 153, Miser and Quade, 1985).

A variety of techniques have been developed for problem formulation, including the Why-What's Stopping Technique (Basadur, Ellspermann and Evans, 1994), mind mapping (Chapter 12 of Wilson, 1993), cause and effect diagrams (Chapter 13 of Wilson, 1993), the Five W's and H Technique (Chapter 3 of VanGundy, 1988) and Dimensional Analysis (Chapter 3 of VanGundy, 1988). All of these techniques basically represent ways of structuring the brainstorming process. In addition to the references listed above, see Chapter 5 of Miser and Quade (1985) and Chapters 2, 3, and 4 of Schoennauer (1981) for further discussion of these and other problem formulation techniques. Further discussion of the Why-What's Stopping Technique is given below.

The Why-What's Stopping Technique

As noted earlier, no problem exists in isolation but is actually a part of a system of problems, called a problematique by the French or a "mess" by Ackoff (1974). In order to be sure that he/she is attacking the right problem, and in order to generate good initial sets of objectives/attributes, alternatives, etc., the analyst and decision makers should be aware of this system of problems.

The Why-What's Stopping Technique (Basadur, Ellspermann, and Evans, 1994) is one approach for generating this system of problems. As a byproduct, one often also obtains objectives/attributes, initial alternatives and constraints, etc. from the use of the technique. In addition, when there are multiple decision makers for a problem, this technique can be an excellent vehicle for communication among the decision makers.

The why-what's stopping technique involves a facilitator and a group of decision makers and stakeholders meeting for a brainstorming session. The output of the session is a "network" of related problems, which will implicitly contain a set of objectives/attributes, preliminary alternatives, constraints, etc. By a network, we mean a system of "nodes' (each node being a problem statement) connected with lines. A line connecting two nodes implies a direct connection between those two problem statements.

The session commences, following an explanation of the technique by the facilitator, with a brief statement of what is thought to be the problem by one of the participants (stakeholder/decision maker). The facilitator will write that statement on a blackboard (or some other device), and ask the group the answer the question:

"Why is this (the initially stated problem) a problem?"

After receiving an answer from one of the participants, the facilitator might continue to ask the question:

"Why else is this a problem?"

referring to the original problem statement.

What would be obtained following this would be a portion of a problem network as shown in Figure 1, where PS1 was the initial problem statement, and PS2 was an answer to the question: "Why is this a problem?" and PS3 and PS4 are answers to the questions "Why else?"

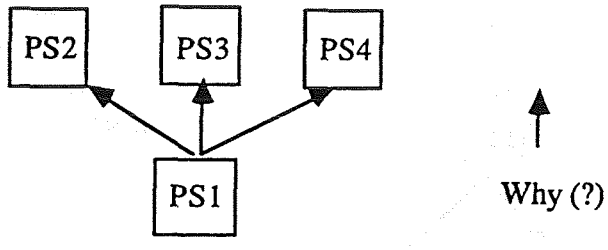


Figure 1. A Portion of a Generic Problem Network

At this point, the facilitator (with input from the participants) could continue to ask "why else" questions in relation to the initial problem statement (PS1), or he/she could select one of the other problem statements (e.g., PS3) and commence the why-why else questioning process all over again. Hence, the facilitator might have a network like the one in Figure 2.

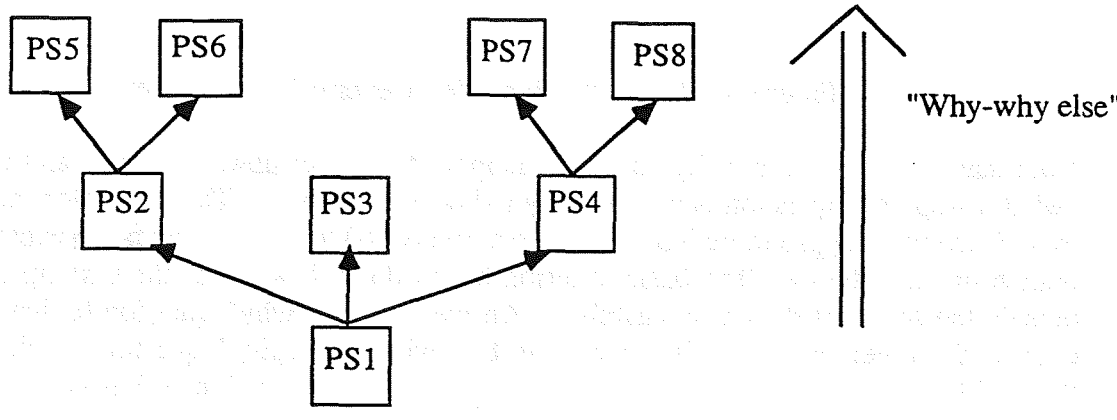


Figure 2. An Expanded Portion of a Generic Problem Network

After a "sufficient" expansion in the upwards direction through the "Why-Why else" questioning process, the facilitator may wish to expand the network in the other direction. This can be accomplished by asking the question:

"What's Stopping us from solving this problem (e.g., problem PS1)?"

This question is usually asked in relation to the initial problem statement (PS1), but it could be asked in relation to any of the other problem statements.

Following this process, one should have a problem statement network which would appear something like the one in Figure 3.

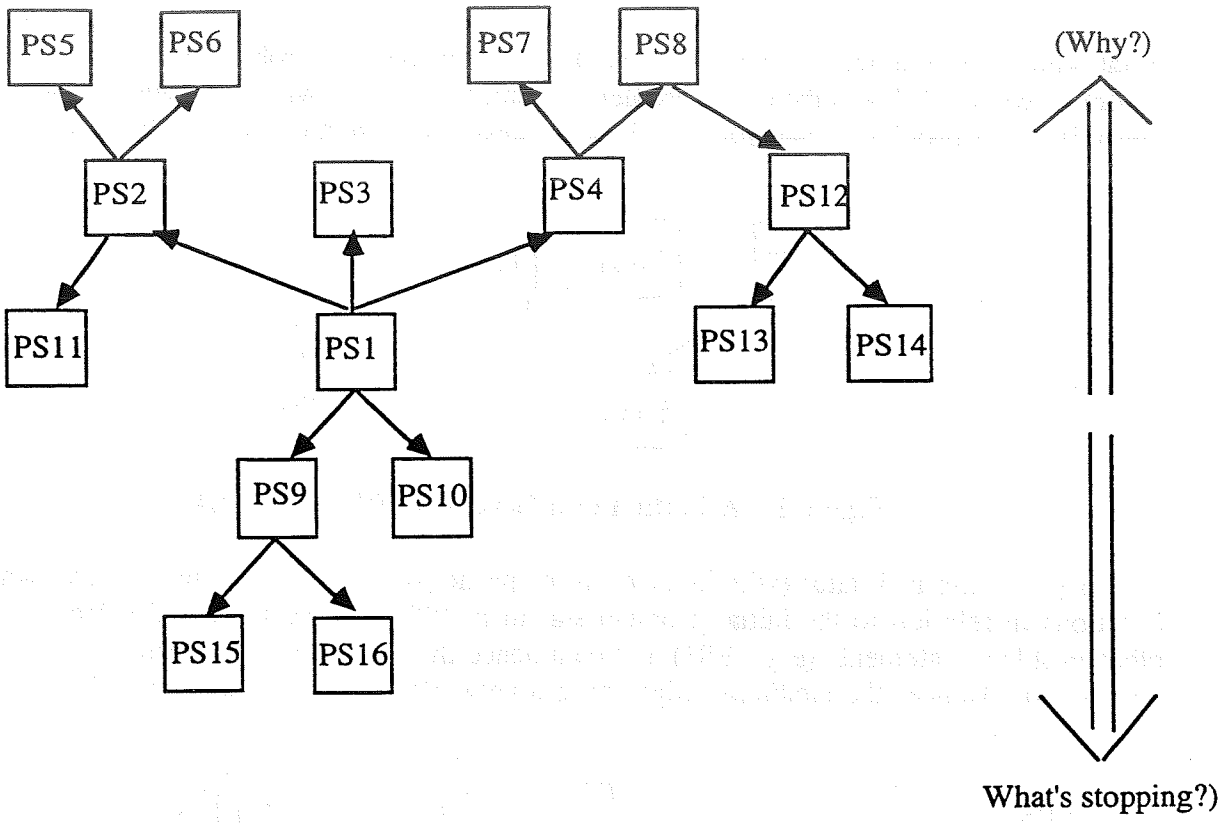


Figure 3. A Totally Expanded Generic Problem Network

Note that an answer to a "why" question is indicated by an upward arrow, and an answer to a "what's stopping" question is indicated by a downward arrow. The facilitator must use his/her experience (with appropriate input from the participants) to decide how far to expand a particular branch of the network. The basic criterion has to do with whether the participants are getting outside the scope of their responsibility. Answers to the "why" question basically result in an expanded problem scope, while answers to the "what's stopping" question result in condensing the problem area, even leading to specific alternatives for problem solution.

The exercise also leads to the realization that there is a fine line between alternative solutions and objective/attribute measures. For example, an answer to the question of

"What's stopping us from improving quality?"

might be:

"A lack of well-trained inspectors."

This leads one to suspect that an alternative solution might be:

"Get more well-trained inspectors."

In addition, in a different, but related, problem context, the number of well-trained inspectors might be an attribute, associated with the problem of developing a plan for increasing the number of well-trained inspectors.

One other activity which the facilitator does is reframe the problem into an "opportunity" or at least a more optimistic statement, by attaching the phrase: "How might we" at the beginning of each problem statement. For example, instead of writing:

"A lack of well-trained inspectors," the facilitator would write:

"How might we increase the number of well-trained inspectors?"

on the blackboard.

Structuring Objectives and Attributes

Objectives and attributes for a problem typically fall into a "hierarchy". At the top of the hierarchy are the more general objectives, while towards the bottom of the hierarchy are the more specific objectives. Finally at the bottom of the hierarchy are the specific attributes, each one connected to a specific objective. As an example, consider the hierarchy below, reprinted from page 105 of French (1986).

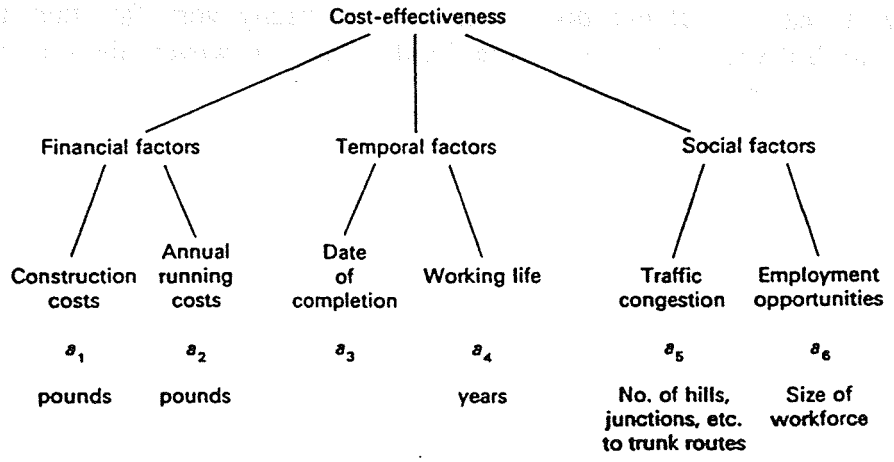


Figure 4. A Hierarchy of Objectives

Note that at the top of this hierarchy is the overall objective: optimize cost-effectiveness. This overall objective is achieved by optimizing financial factors, temporal factors, and social factors. Optimizing financial factors means minimizing construction costs and minimizing annual running costs. Attributes are used to measure the lowest level objectives. For example, the objective of minimization of construction cost is measured by construction costs in dollars. Construction of a hierarchy is an important aspect of problem definition. A set of lower level objectives can be attained from a higher level objective by asking a question such as:

How do we achieve (the higher level objective)?

For example, the way to optimize financial factors is to minimize construction costs and to minimize the annual running costs.

The question of how far to extend a hierarchy for a particular problem is subjective in nature, relating to several different aspects of the problem (e.g., the orientation of the decision maker--quantitative or nonquantitative; the nature of the evaluative models to be used; etc.). If one does not extend the hierarchy very far then, typically, the attributes that would be used to measure the associated higher level objectives would be subjective in nature. See Chapter 2 of Keeney and Raiffa (1993) for further discussion on construction of a hierarchy.

Attributes associated with the lowest level objectives of a hierarchy may be either quantitative or nonquantitative (i.e., subjective) in nature. Quantitative attributes (e.g., construction cost in current dollars) can typically used when the hierarchy is extended to include very specific objectives. Nonquantitative, or subjective, attributes are used when the hierarchy is not extended very far; i.e., when the lowest level objectives of the hierarchy are subjective in nature.

Obviously, there are tradeoffs associated with the decision of how far to extend a hierarchy. For example, if one does not extend the hierarchy very far, one will have the advantage of having to deal with only a very few attributes. However, these attributes will probably be subjective in nature. If one does extend the hierarchy very far, then there may be many attributes to deal with in the analysis (a disadvantage); however, these attributes will probably be quantitative in nature.

4. MODELING AND SOLUTION METHODOLOGIES

As discussed earlier, there are two types of models and associated solution methodologies required for solving a well-constructed problem: evaluation models and associated solution methodologies, and ranking/optimization models and associated solution methodologies. This section of the report discusses several types of these models and solution methodologies, as well as associated software packages.

A model refers to a representation of a system and/or a decision maker's preference structure over the outcome space, while a solution methodology refers to how you experiment with (or solve) the model in order to select a best alternative solution to the problem.

Evaluation models provide a mapping from the alternative space to the outcome space (as measured by the attribute values). For example, in a problem situation involving a modification to orbiter processing procedures, the two alternatives to consider would be:

1. Do not implement the modification.
2. Implement the modification.

The objective that might be considered here would be:

Maximize the net present worth of expected savings.

Hence, one would need some type of evaluation model (represented as a sequence of equations) which would output the net present worth of expected savings as a function of the proposed modification.

The various types of models, methodologies, and techniques from operation research cannot be easily classified as either evaluative in nature or optimization-based in nature; often, a particular type of model allows the performance of both functions. However, the emphasis is usually on one area or the other.

For the purpose of this report, we will classify the following types of models as emphasizing the evaluative process: simulation models, analytic models, gaming models, judgmental models, decision trees, influence diagrams, and decision tables. These model types are discussed below.

Simulation models are useful evaluation tools when a detailed representation of a process in required. As defined by Pegden, Shannon, and Sadowski (1990), simulation is "the process of designing a model of a real system and conducting experiments with this model for the purpose of understanding the behavior of the system and/or evaluating various strategies for the operation of the system." Simulation models are characterized by the fact that they contain no inherent sophisticated optimization capability, and that their outputs are typically multiattributed, probabilistic outcomes.

Usually, several different statistical problem must be addressed when one is using a simulation model. One reason for this is that one run of a simulation model corresponds to a statistical experiment. Examples of statistical problems that one must address when using a simulation model include input modeling (fitting distribution functions to data), generating samples from probability distributions, setting initial conditions for the simulation runs, determining a "warm-up period" for the run, and various problems related to interpretation of output--estimating expected values and variances, determining confidence intervals, hypothesis tests, and ranking probabilistic outcomes with a specified degree of confidence. See Law and Kelton (1992) for further discussion of these and other statistical concepts.

A variety of software packages have been developed to aid in the simulation modeling process. In fact, the number of these packages has grown tremendously in the last 10 years, indicating the popularity of this modeling methodology. A good source for detailed descriptions of many of these software packages is the Software/Modelware Tutorial Section of the annual Proceedings of the Winter Simulation Conference.

A few years ago, one could classify these simulation software packages as being either simulators (in which no programming was required to build a model) or simulation languages. Recently however the distinction between these two categories has blurred. For example, the Arena package of System Modeling Corporation allows programming inserts which make it extremely flexible. Examples of some of the other simulation software packages include SIMAN/Cinema, SLAMSYSTEM, GPSS, SIMSCRIPT, WITNESS, PROMOD, MODSIM, Taylor II, etc. See Pritsker (1995) for additional discussion on simulation.

In an analytic model, mathematical statements are used to represent the relation that hold between the variables of interest (page 195, Miser and Quade, 1985). These models typically do not represent a system in as much detail as a simulation model, hence they are typically not as accurate as simulation models. Instead, analytic models represent a system as a group of (either simultaneous, or serial) equations.

In addition to consisting of either a simultaneous or serial system of equations, analytic models can be either deterministic or probabilistic in nature. If the system is probabilistic in nature, then typically one may need to employ some type of sampling procedure in the solution process. The Demos software package (Lumina Decision Systems, 1993) is an example of a package which will solve a probabilistic analytic model consisting of a series of equations.

Basically, whenever the relationship(s) between a set of alternatives and the chosen attributes/objectives can be represented as closed form equations, then an analytic model can be used. If the relationship(s) cannot be represented in this fashion, or are not even known, then one may have to employ a simulation model.

Analytic models can be useful in the calculation of economic performance measures such as net present worth, or in solving inventory control problems, queueing problems, and network flow problems. Mathematical programs are basically analytic models.

A gaming model is a form of simulation modeling where the people involved with the system under study simulate the behavior of major elements of the system. Typically, these models require much effort to develop since they require interaction between humans and computers. Gaming models, which have been used extensively by the military establishment, are quite useful as training tools and to improve communication among players of different disciplines. See pages 197 to 199 of Miser and Quade (1985) for further discussion of gaming models.

Judgmental models basically involve gathering a group of experts to forecast the outcomes associated with various alternative problem solutions. These models can be useful when there is little data available for building simulation or analytic models. The various types of judgmental models differ in the way they structure the group discussion. Examples of techniques associated with judgmental modeling include scenario writing (see Section 9.5 of Miser and Quade, 1985), the Delphi technique (Linstone and Turoff, 1975), cross-impact analysis, and various team and workshop approaches.

In problem situations where the alternatives are expressed in terms of sequences of interrelated decisions, a decision tree may be an appropriate modeling technique. These decision trees explicitly account for uncertain factors in the decision making process, as illustrated by Figure 5, reprinted from page 260 of French (1986).

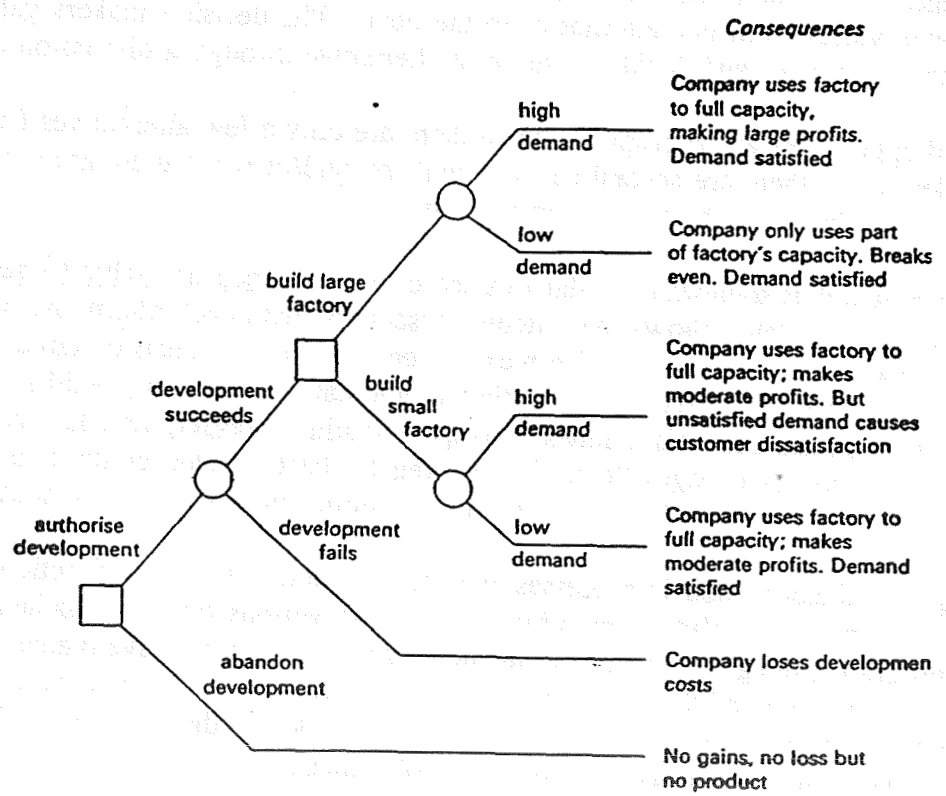


Figure 5. An Example of a Decision Tree (page 260, French, 1986)

In this problem, the initial decision to be made (authorize or abandon development) is dependent upon the outcomes associated with later decisions. In addition, the problem includes probabilistic events (e.g., development succeeds or development fails) over which the decision maker has no control. A decision tree analysis would indicate which decision is the best to take initially.

An influence diagram might be thought of as a generalization of a decision tree, in which the influences which affect variables are indicated through a graph of nodes and arcs. The nodes represent the variables and an arc leading from one node to another indicates that the first variable has an affect on the second. The software package, Demos, utilizes the concept of influence diagrams. Other software packages which model decision trees and/or influence diagrams include DPL, Arborist, and Supertree. See recent issues of the journal OR/MS Today for reviews of these and other software packages for decision analysis. In particular, see the article by Shachter (1986) or Bunn (1984) for further information on influence diagrams.

The scorecard approach to problem solving (see pages 231 to 236 of Miser and Quade, 1985) is fairly informal in nature. In this approach, all of the consequences associated with an alternative (quantitative or qualitative in nature) are displayed in a tabular array. Probabilistic consequences are also displayed. In the tabular array, the entries in each column represent the consequences associated with an alternative, while the row entries show how a particular consequence varies from one alternative to the next. The decision makers gather in a room, develop the scorecard, and decide on the best alternative through a discussion process.

Scorecard approaches are appropriate when there are only a few alternatives (say 10 or fewer) to consider, when there are several decision makers, and/or when there are many attributes, at least some of which may be probabilistic in nature.

A construct which is somewhat similar to a scorecard is the decision table (page 163 of French, 1986). A decision table shows the outcome associated with each alternative for each state of nature. Once the alternatives have been expressed in terms of a decision table, one can employ any of several different criteria for selecting an alternative, including Wald's maximim return, Hurwicz's optimism-pessimism index, Savage's minimax regret, and Laplace's principle of insufficient reason (see pages 36 to 39 of French, 1986). One could also employ a more sophisticated approach, involving a multiattribute utility function, to be discussed later.

Ranking/optimization models/processes include multiattribute value functions, multiattribute utility functions, the analytic hierarchy process, and various types of mathematical programs (linear programs, integer programs, dynamic programs, multiobjective mathematical programs, etc.). The first three of these: multiattribute value functions, multiattribute utility functions, and the analytic hierarchy process concentrate on modeling the decision maker(s) preference structure over a multidimensional, possibly probabilistic, outcome space.

A multiattribute value (MAV) function is a function which maps scores over multiple attributes into an interval of the real number line, usually [0,1]:

$$v: x_1, x_2, \ldots x_n \rightarrow [0,1]$$

where $x_1, x_2, \ldots, x_n$ are the n attribute values. A MAV function has the property that if the decision maker prefers one outcome to another, in terms of its multiple attribute values, then the MAV function value associated with the first outcome will be higher than the value associated with the second outcome. In symbolic terms:

$$(x_1', x_2', \ldots, x_n') \succ (x_1'', x_2'', \ldots, x_n'')$$

if and only if

$$v(x_1', x_2', \ldots, x_n') > v(x_1'', x_2'', \ldots, x_n'')$$

where the symbol "$\succ$" means "is preferred to."

Once a MAV function has been developed, instead of optimizing over n attributes, one only has to optimize over one measure.

MAV functions are formulated through interview sessions between the analyst and the decision makers, in which the decision makers must answer hypothetical questions concerning his/her tradeoffs over multiple, conflicting objectives. Note that two different rational and reasonable decision makers could have different MAV functions, depending on how they value the different objectives. See Chapter 3 of Keeney and Raiffa (1993) for more information on MAV functions.

A multiattribute utility (MAU) function represents a generalization of a MAV function in that it allows a ranking of alternatives in situations with multiple objectives and uncertainty/risk. More formally, a MAU function is a mapping from an n-dimensional attribute space: $x = (x_1, \ldots, x_n)$ to a closed interval: [0, 1] of the real number line. It has the property that:

$$\tilde{x}' \succ \tilde{x}'' \text{ if and only if } Eu(\tilde{x}') > Eu(\tilde{x}'')$$

where $\tilde{x}'$ and $\tilde{x}''$ are two different n dimensional probabilistic outcomes and $Eu(\tilde{x}')$ and $Eu(\tilde{x}'')$ are the expected utilities associated with those outcomes, respectively. The above basically says that the decision maker prefers outcome $\tilde{x}'$ to outcome $\tilde{x}''$ if and only if his/her expected utility for outcome $\tilde{x}'$ is larger than the expected utility for outcome $\tilde{x}''$.

Assessment of a MAU function is usually much more difficult than the assessment of a MAV function since the decision maker must answer questions about hypothetical situations involving multiple objectives and uncertainty/risk. See Keeney and Raiffa (1993) or French (1986) for

further information about MAU functions.

A MAU function can be an excellent device for ranking outcomes involving multiple objectives and uncertainty/risk. They can be used in conjunction with other methodologies such as simulation, decision trees, influence diagrams, and decision tables. In addition, just the assessment process by itself can be useful in helping the decision makers to think about their preferences in a structured fashion.

The analytic hierarchy process (AHP) relies on the development of a hierarchy of important problem factors (e.g., objectives or attributes) to rank alternatives over multiple attributes. In the hierarchy, the higher level objectives are at the top of the hierarchy, while the lower level objectives are towards the bottom. The second lowest level of the hierarchy will typically consist of problem attributes while the lowest level will consist of the various alternatives. Through sets of pairwise comparisons of the factors at each level of the hierarchy, local and global weights, reflecting importance, are computed for each element of the hierarchy. The global weights computed for the respective alternatives are the scores for those alternatives--the higher an alternative's score, the more desirable that alternative is.

AHP is implemented in a software package called Expert Choice, advertised in OR/MS Today. The score by AHP for an alternative is comparable to the score that the alternative would receive through the use of a MAV function. AHP is typically easier to use for a ranking process than a MAV function, because of the difficult assessment process associated with a MAV function. However, there are theoretical problem associated with AHP, as reported in the literature (e.g., see various issues of Management Science). One of the most basic problems with AHP is that the decision maker must answer questions such as:

"How much more important is safety than cost?"

without the use of any type of scale. A MAV function assessment on the other hand involves asking the decision maker detailed questions concerning tradeoffs between specific outcomes, e.g.:

> "How much would you be willing to pay in order to decrease the probability of serious injury in a particular year from .00001 to .000009?"

In summary, if an accurate representation of the decision maker's preference structure is needed, then a MAV function in probably more appropriate to use than AHP. If one needs a quick answer to a problem, than AHP is probably more appropriate. See Saaty (1988) for further discussion of AHP.

None of the three approaches discussed above (MAV functions, MAU function, or AHP) include any inherent optimization capability. When there are a small number of alternatives to consider (e.g., 10 or fewer) no sophisticated optimization capability is needed, and a "complete enumeration" of all of the alternatives can be performed. When there is a large number or even

an infinite number of alternatives, some type of implicit enumeration is needed. This turns out to be the case when one has combinations of things to consider or there are continuous decision variables to consider. Examples of the first situation (combination) would be the following:

1. A committee has 100 different (but dependent) processing enhancement suggestions to consider. Any number and combination of suggestions can be implemented; hence, the number of alternatives is actually the number of combinations (2^{100}) to consider.
2. Shuttle processing requires literally thousands of activities. The timings (when they are performed) of these activities are important decisions to be made. The timing/sequence of activities is restricted by precedence relationships (e.g., it's necessary to inspect some item before it can be replaced/repaired), resource constraints (e.g., the number of technicians available is limited), or capital constraints, among other types of restrictions. However, at any particular point in time (e.g., 7 am on August 7) there still may be dozens of activities from which to choose. The various alternatives in this case correspond to the combinations of activities that can be performed at any particular point in time. Note that this is an ongoing decision problem for any large project, and that this area of research (resource-constrained project scheduling) has been addressed extensively in the literature.

An example of a continuous decision variable would be the determination of the length of time to test some item. Since this length of time is continuous, the number of alternatives is, in effect, infinite.

A wide variety of mathematical programming models allows one to represent optimization situations like the ones described above. Examples of these mathematical programs include linear programs, integer programs, dynamic programs, goal programs, and multiobjective mathematical programs. See Nemhauser (1989) for further discussion of some of these models.

A survey of the literature reveals the existence of a large number of sophisticated optimization techniques for solving optimization problems. These techniques can be categorized as being either heuristic or exact in nature.

Heuristic optimization techniques are those which give a good, though not necessarily optimal, solution to a problem. One advantage of these techniques is that they are usually easy to understand and explain to others. They do not usually require any type of sophisticated mathematical formulation of the decision problem, and can usually be easily implemented on a computer. One example of a heuristic optimization technique would be to evaluate 1000 alternative, randomly chosen, solutions to a problem (out of say approximately one million feasible alternative solutions), and then select the best one out of the 1000 to implement.

Resource-constrained project scheduling is an example of an area which relies heavily on heuristic optimization techniques. Many of these techniques have been implemented in several

different software packages for project scheduling.

Numerical search procedures for solving nonlinear optimization problems is another example of a body of techniques that can be thought of as being heuristic in nature. These procedures typically have somewhat subjective termination criteria, and in addition, usually achieve, at best, a local optimum to a problem.

Examples of exact optimization techniques include those often associated with mathematical programming models: linear programs (revised simplex method, interior point methods), integer programs (branch-and-bound methods, dynamic programming), nonlinear programs (methods based on differential calculus, dynamic programming). These complex techniques are often difficult to explain to others, and often require the formulation of the problem into one of the mathematical constructs listed above. Sometimes exact techniques can be used in a heuristic fashion; for example, one has the option in a branch-and-bound procedure (for solving an integer program) to terminate the procedure when the algorithm has achieved a solution that is within x % of an optimal solution in terms of the objective function value.

The choice of which optimization technique to choose for situations where there are a large number of alternative is, of course, somewhat subjective in nature. When the evaluation model is very complex (e.g., highly nonlinear or not even of a closed-form), then one would probably want to use a heuristic optimization technique. A situation involving the use of a simulation as an evaluation model is one where the model would not be of a closed form, and hence one may want to use a heuristic technique.

When the evaluation model can be formulated as a mathematical program, then the possibility of using an exact optimization technique exists. Typically, however, such an optimization technique can only be implemented through the use of an "expensive" software package, such as CPLEX for solving linear program or integer programs. See various issues of the journal OR/MS Today for description of the various software packages available for the optimization process. See Nemhauser (1989) for further discussion on optimization techniques.

Several of the models and associated solution methodologies can be incorporated into a single decision support system for a particular problem situation. For example, it may be reasonable to combine a simulation model with a utility function and a numerical optimization technique in order to select a best solution to a problem with multiple objective and uncertain outcomes. The best techniques to use for a particular problem depend upon the characteristics of that problem, discussed in the next section of this report.

5. PROBLEM CHARACTERISTICS AND THE CHOICE OF APPROPRIATE SOLUTION METHODOLOGIES

In this section of the paper we briefly discuss the important characteristics of a problem which affect the methodologies chosen to solve that problem.

Problem importance is probably the major characteristic associated with the choice of techniques/methodologies for solving a problem. Importance can be measured in terms of performance measures related to cost, schedule, safety, etc. or in terms of political pressures (e.g., who in the organization is interested in the problem). Obviously the amount of time and money that one will spend in solving a problem (and hence, the solution methodologies chosen) should depend on problem importance.

The number and types decision makers involved in a problem is also an important problem characteristic. The predisposition of the decision maker(s) towards the use of quantitative techniques is an important factor to consider, especially as it relates to the choice of a preference structure model. (Note that while it is important for the decision maker to be involved in building both the evaluation models and the preference structure models, this involvement is more crucial in terms of the preference structure models). Also, the larger the number of decision makers involved in a problem, the more difficult that situation. Most methodologies/techniques for development and analysis of preference structure models make the assumption that there is only <u>one</u> decision maker. For example, the procedure for formation of a multiattribute value or utility function assumes that there is <u>one</u> decision maker answering question concerning tradeoffs among performance measure. If there were multiple decision makers, an analyst would basically have two choices:

1. Assemble the decision makers as a group and, for each question regarding hypothetical tradeoffs, have them provide a consensus answer.
2. Assess a value/utility function for each decision maker, and then rank the alternatives for each decision maker.

In the second case, the decision makers would meet as a group and discuss their individual rankings in order to arrive at a consensus first choice alternative. Hopefully, the individual ranking would not differ from each other too much.

Of the two approaches listed, the first one would be the preferred approach with respect to time and effort, especially if there were not too many decision makers.

The number of attributes associated with the problem affects both the type of evaluation model(s) and the type of ranking/optimization model(s) used for problem formulation. Note that often an analyst will have much leeway with respect to the number of attributes used in the analysis, depending upon how far down the hierarchy of objectives/attributes he wishes to move. In cases where only a very few attributes are used in the analysis (say, one to three), the attributes themselves may be very subjective in nature. Hence, it may be difficult to get a good,

"quantitative," evaluation model; what may be used instead is a judgmental model to relate alternatives to attribute values.

An example of this situation may be one where a department must rate Small Business Innovation Research (SIR) Program proposals, over four subjective attributes: scientific/technical merit, anticipated commercial applications, qualification, and merit of proposed work plan. In this case, the alternatives are the individual proposals, respectively, and the evaluators, through their mental models, give scores to each of the proposals over each of the attributes.

Of course, there may also be situations with a very few attributes where these attributes are quantitative in nature. In these situations, more sophisticated quantitative models (e.g., simulation or analytic models) may be used.

Also, in situations where there are very few attributes, the preference structure model used can typically be very sophisticated (e.g., a multiattribute utility or value function). The reason for this is that the amount of "preference" information required to assess a multiattribute utility or value function becomes onerous as the number of attributes becomes large.

In situations where the analyst decides to use a large number of attributes (either because the problem situation calls for it, or for more arbitrary reasons), the attributes themselves may very well be quantitative in nature; hence, a sophisticated, and explicit, quantitative evaluation model may be appropriate. However, as discussed above, a sophisticated preference structure model (e.g., a multiattribute value or utility function) would be very difficult to develop in situations involving a large number of attributes. In this case, therefore, the decision maker may want to use a scorecard approach (Miser and Quade, 1985) or a constraint-based approach, as discussed below.

There are many different constraint-based procedures that one could employ when there are many attributes to consider. For example, the analyst might assess a utility/value function over one or two of the more important attributes, and then constrain the other attribute values to attain particular levels. The approach may be "interactive" in nature in that the decision maker(s) may vary the goal levels, similar to one of the approaches involving multiobjective mathematical programming.

The complexity of the relationships between the alternatives and the outcomes and the amount of hard data available to assess these relationships affects the choice of the type of evaluation model used. For example, if there is little or no data available, and the relationship is a complex one, then judgmental model(s) may have to be used. This may be the case if the alternative being evaluated relates to a completely new type of system.

If the relationship(s) are complex, yet the processes which give the relationships are well-understood, then a simulation model may be appropriate to use as an evaluation tool. If the relationship(s) are well-understood and can be written as a series of equations, than an analytic

model may be appropriate.

The amount of uncertainty/risk involved in the relationship between the alternatives and the outcomes can affect the type of evaluation model used and the methods used to experiment with the model. For example, simulation models are excellent choices for representing the uncertainty in a system. However, if a simulation model cannot be used, then sensitivity analysis of an analytic model may be appropriate for analyzing how changes in inputs can affect the outputs associated with a system.

The number of alternatives under investigation affects the choice of a ranking/optimization model and associated solution technique. For example, when there are only a very few alternatives to consider (e.g., ten or fewer) then a complete enumeration of all alternatives can be performed -- no sophisticated optimization procedure is needed. In cases with many alternatives (as a result of continuous decision variables or combinations of decisions) some type of efficient representation of the alternatives is needed, as well as a sophisticated optimization technique.

There are several other considerations that could affect the choice of modeling techniques. In particular, the time available to solve a problem affects the choice, as well as whether or not the problem situation is an ongoing one. For example,if one does not have a lot of time to solve the problem and there is little or no data available to help in the development of an evaluation model, than a judgmental model might be appropriate. As another example, if the problem situation is ongoing (e.g., scheduling and allocation of resources for orbiter maintenance), then a decision support system, which allows for collection of data to continually update a data base, may be an appropriate approach.

6. APPLICATIONS AT THE KENNEDY SPACE CENTER

The techniques of decision analysis, and related methodologies, can be extremely useful tools to the Industrial Engineering Staff at KSC. Some example applications that come to mind include:

1. Scheduling/allocation of resources for orbiter maintenance and repair.
2. Development of an inventory policy for important spare parts of the orbiter.
3. Investigation of the affects associated with retraining of technicians to allow greater sharing of resources among orbiter bays.
4. Investigation of flow time change as a result of a major change in the orbiter.
5. Investigation of the relationship between resource levels (e.g., number of technicians) and orbiter flow time.

In this section of the report, we will describe four specific applications of some of the techniques discussed earlier to problems arising at the Kennedy Space Center. The emphasis in these applications is on the consideration of probabilistic aspects of the problem, through the use of the Demos software package. A separate report has been submitted on the use of the Demos software package (Evans, 1995).

Evaluation and Ranking of Phase I Small Business Innovation Research Program Project Proposals

Each year, members of the Industrial Engineering Group at KSC evaluate Phase I Small Business Innovation Research (SBIR) program proposals. As specified in the Phase I Evaluator Handbook, the rankings are accomplished by scoring the proposals in each of four different criteria:

1. Scientific/Technical Merit.
2. Anticipated Commercial Application of the Technology.
3. Qualifications (of the key project personnel).
4. Merit of the Proposed Work Plan.

Each proposal is given a score of 0 to 40 points on the first criterion, 0 to 25 points on the second criterion, 0 to 25 points on the third criterion, and 0 to 10 points on the fourth criterion. The total score for a proposal is the sum of its scores on the four criteria.

The Industrial Engineering Group desired a procedure for scoring that would allow probabilistic inputs for each of the four criteria. For example, an evaluator may be very uncertain about the anticipated commercial applications associated with a particular proposal. He/she may think that the potential might be excellent or might be only fair.

After consultation with the IE Group, it was decided that an approach involving the use of a multiattribute utility function would be appropriate for this application. In addition, it was

determined that expansion of the hierarchy past the four criteria listed above would not be desirable. Hence, since these criteria are somewhat subjective in nature, subjective attributes were used in the evaluation process. These subjective attributes assign numerical ratings to various attribute levels, as follows:

10: excellent,
8: very good,
6: good,
4: fair,
2: poor,
0: very poor.

Intermediate scores were also allowed. For example, a score of 8.5 is one-quarter of the way from "very good" to "excellent."

Also, since the weightings had been established as 40, 25, and 10 points, respectively; and since it was specified that a proposal's total score was to be computed as the sum of its scores on the four criteria, an additive utility function of the following form was chosen for this application:

$$u(x_1, x_2, x_3, x_4) = .4u_1(x_1) + .25u_2(x_2) + .25u_3(x_3) + .1u_4(x_4)$$

where $u_i(x_i)$ represents the i^{th} individual attribute utility function and x_i is the (possibly probabilistic) score for the i^{th} attribute. (See Keeney and Raiffa (1993) for discussion of additive utility functions and individual attribute utility functions).

As noted by Keeney and Raiffa (1993) and others, an additive utility function is very restrictive with respect to the decision maker's preference structure. However, given the restrictions imposed by the Evaluator Handbook (that the total score be the sum of the individual scores), the additive utility function was an appropriate choice.

The utility function approach yielded two important advantages. First, probabilistic (or uncertain) scores could be entered for each of the four criteria. Second, the risk preferences of the evaluators could be explicitly considered. For example, an evaluator may prefer a proposal which has a score of 8 (very good) on technical merit for certain, to a proposal which has a 50% chance of a score of 6 and a 50% chance of a score of 10, even though the expected scores on each proposal are the same.

The individual attribute utility functions were then assessed through an interview process with the staff of the Industrial Engineering Department. This process employed the midvalue splitting technique as described by Keeney and Raiffa (1993). The output from the process was a series of points on the graph of the individual attribute utility functions. The linear interpolation feature of the Demos package was employed to compute the utility function values associated with attribute scores. These coordinates on the graphs are given below:

x_1	0	3.1	3.8	4.5	5.3	6.5	7.7	8.8	10
$u_1(x_1)$	0	.25	.375	.5	.625	.75	.875	.9375	1

x_2	0	3.7	5.8	6.8	8	9	9.5	10
$u_2(x_2)$	0	.25	.5	.625	.75	.875	.9375	1

x_3	0	4.5	5	7	8	10
$u_3(x_3)$	0	.25	.5	.75	.875	1

x_4	0	3.4	5.15	6.05	7.25	8.35	9.15	10
$u_4(x_4)$	0	.25	.5	.625	.75	.875	.9375	1

where x_1, x_2, x_3, and x_4 refer to the scores attained on the four criteria: Scientific/Technical Merit, Anticipated Commercial Application of the Technology, Qualifications, and Merit of the Proposed Work Plan, respectively.

These utility function values were input to a Demos model for computation of expected utility. The model is currently set up to accept probabilistic (or deterministic) scores for a proposal on each of the four criteria for a particular proposal by a particular evaluator. Specifically, the model accepts parameter values for a triangular distribution over each of the four criteria:

a: worst possible score
m: most likely score
b: best possible score

If the evaluator decided that he/she were certain about a score on a particular proposal for a particular criterion, then the three inputs (a, m, and b) would have the same values. In addition, if an evaluator wished to use a different distribution to represent his/her uncertainty about a proposal's score on a particular criterion, then one of the other standard distribution functions from Demos could be employed with little change to the current model.

Output from the Demos model includes expected weighted utilities over each of the four criteria as well as expected total utility. The Demos diagram for the model is shown in Figure 6.

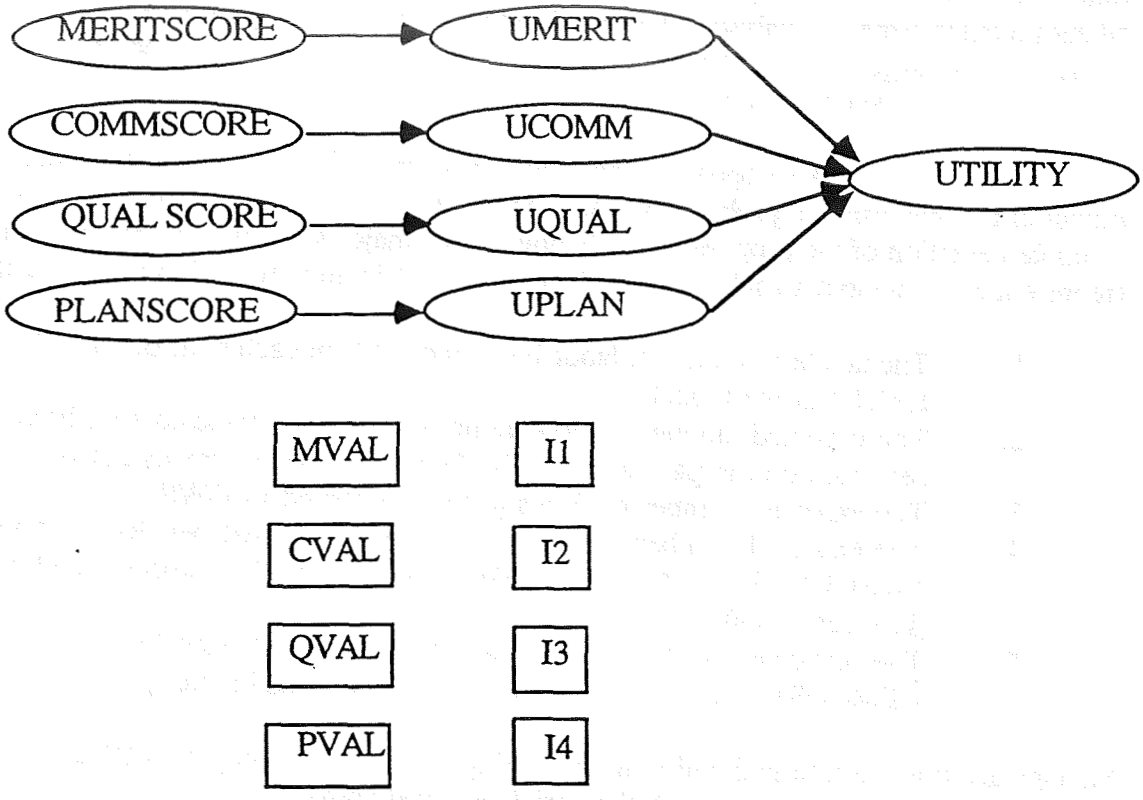


Figure 6. Demos Diagram for the SBIR Proposal Ranking Model.

In this model, the probabilistic scores for the four criteria are input through the appropriate expressions for the nodes MERITSCORE, COMMSCORE, QUALSCORE, and PLANSCORE, respectively. The points on the individual attribute utility function are input through the nodes MVAL, I1, CVAL, I2, QVAL, I3, PVAL, I4.

OMS Heater Mod Benefits Analysis

In 1990 a suggestion was made concerning a modification to orbiter processing activities. By incorporating a simple wiring modification to each orbiter, the necessity to remove the 59-63 and 59-64 doors for the Orbiter Maneuvering System (OMS) Pod, Heater checks could be eliminated. Currently, door removal subjects tiles, filler bar, and associated hardware to potential damage. Removals, inspection, reinstallation, and associated PR/DRs (e.g., as a result of TPS repairs) results in several hundred manhours of direct labor effort. In addition, if the doors are removed on the pad, the possibility of impacting the critical path exists, which can result in overtime and other potential costs.

A various of analyses have been performed since 1990. Finally, in 1995 a group from the Industrial Engineering Department at the University of Central Florida submitted a "deterministic" benefits analysis. The KSC Industrial Engineering Department thought that, as a result of the large number of uncertain variables associated with the problem, a probabilistic benefits analysis would be appropriate.

A Demos model was developed to implement this probabilistic benefits analysis. The main output of this analysis is a probability distribution over expected cost savings per year as a result of implementation of the proposed modification. The major sources of uncertainty in the model (represented as random variables in the Demos model) include the following estimates:

1. The number of indirect labor hours required for each hour of direct labor (labeled RATIO in the model).
2. The expected number of door removals per flow (labeled PDRFLOW for door removals on the pad, and NPDRFLOW for door removals not on the pad).
3. The expected number of flows per year (labeled FLOWS).
4. The expected number of direct labor hours required per door removal (labeled PDHOURDR and NPDHOURDR, respectively, for expected pad and non-pad door removals).
5. The expected material cost per door removal (labeled PMCOSTDR and NPMCOSTDR, respectively, for pad and non-pad material costs).

Savings are divided into pad and non-pad savings. The uncertainty in total savings is a result of the uncertainties associated with the variables listed above.

The diagram corresponding to the Demos model for this application is shown in Figure 7.

An arc pointing from one node to a second node in the diagram indicates a dependence of the second node's variable on the first node's variable. Each node has a specific definition, which defines its relationship to other variables. For example, for the variable NPSAVINGS (expected yearly savings from door removals done off the pad), this definition is given by:

NPSAVINGS = NPDRY * (NPMCOSTDR + NPDCOSTDR + NPICOSTDR)

where,

NPDRY ≡ number of non-pad door removals per year.

NPMCOSTDR ≡ material cost per non-pad door removal.

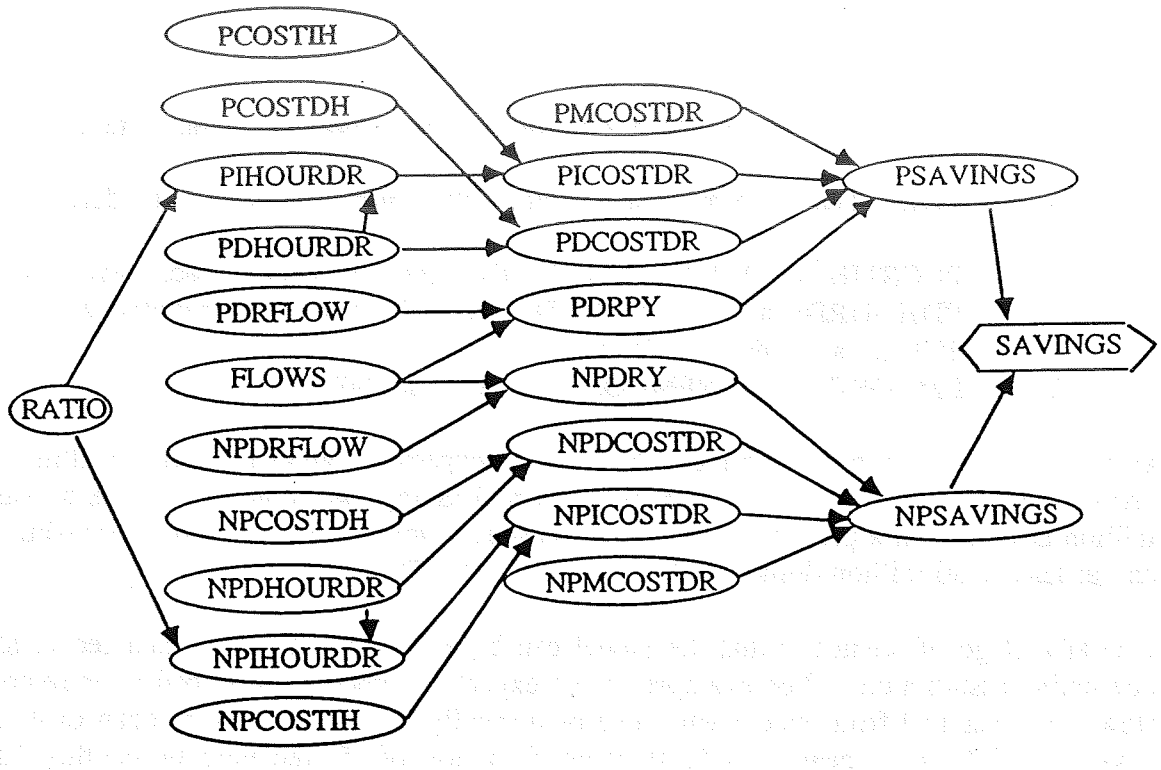


Figure 7. Demos Diagram for the OMS Heater Mod Benefits Analysis Model.

NPDCOSTDR ≡ direct labor cost per non-pad door removal.

NPICOSTDR ≡ indirect labor cost per non-pad door removal.

Other variable definitions in the model are given by:

PCOSTIH (NPCOSTIH) = Pad (non-pad) cost per indirect hour of labor.

PCOSTDH (NPCOSTDH) = Pad (non-pad) cost per direct hour of labor.

PIHOURDR (NPIHOURDR) = Number of indirect hours generated per door removal on the pad (not on the pad).

Assuming the following values of some of the input variables of the model:

1. RATIO is normally distributed with a mean of 4.6 and a standard deviation of 1.
2. FLOWS* has a value of 6 with a probability of .2, 7 with a probability of .6, and 8 with a probability of .2.

*Note that, more realistically, since FLOWS is supposed to represent the expected number of flows in a year, it should be modeled as a continuous rather than a discrete random variable.

3. NPDRFLOW is normally distributed with a mean of 4 and a standard deviation of 1.
4. PDRFLOW is normally distributed with a mean of .5 and a standard deviation of .1.
5. PCOSTIH, PCOSTDH, NPCOSTDH, and NPCOSTIH each have values of 40.
6. PDHOURDR and NPDHOURDR are each normally distributed with a mean of 250 and a standard deviation of 50.
7. PMCOSTDR and NPMCOSTDR are each $5000.

Demos was able to output a sample distribution of expected savings per year resulting from the proposed modification. For example, the expected savings was computed to be at least 2.283 million dollars with a probability of .25, at least 1.84 million dollars with a probability of .5, and at least 1.56 million dollars with a probability of .75.

One advantage of Demos is that the model can be easily, and quickly, changed to allow for alternative assumptions. For example, if the expected number of non-pad door removals per flow were changed from its current value of normally distributed with a mean of 4, standard deviation of 1, to a mean of 1.5, standard deviation of .5, the new probability bands for expected total savings per year would be given by: at least 1.042 million dollars with a probability of .25, at least 840000 dollars with a probability of .5, and at least 597000 dollars with a probability of .75.

Another feature of Demos is its ability to perform an "importance analysis." This allows the model user to understand how much each uncertain input contributes to the uncertainty in the output, by computing the rank order correlation between the sample of output values and the sample for each uncertain input. See pages 4-12 and 4-13 of the Demos User's reference for instructions on how to perform an importance analysis with a Demos model.

A General Approach to Assessment of Orbiter Processing Modifications

It was suggested by the personnel of the KSC Processing and Enhancements Division that a Demos model could be developed to assess the value of orbiter processing modifications in general. This section of the report outlines a general Demos model for this type of applicaiton.

Processing modifications can generally result in changes relating to four different types of criteria: cost/savings, time, safety, and quality. With repect to safety and quality, certain standards must be met--i.e., if a processing modification does not meet these standards for safety and quality, it will not be considered further. Hence, the problem reduces to the consideration of cost/savings and time.

Costs/savings can be a one-time occurrence (e.g., an initial cost for a modification) or ongoing (e.g., relating to a savings in direct labor hours associated with each orbiter flow). Time refers to an increase/decrease in expected flow time for orbiter processing. A decrease in orbiter flow time will occur only if the relevant activity/group of activities are on the critical path for the

entire flow of activites.

Given these assumptions therefore, a utility function can be used to rank proposed orbiter processing modifications. The utlity function would have three attributes:

x_1 = expected yearly cost/savings associated with the proposed modification, in dollars. (In this case, x_1 will be positive for a savings, negative for a cost).

x_2 = expected initial cost/savings associated with the proposed modification, in dollars. (In this case, x_2 will be positive for a cost, negative for a savings).

x_3 = expected increase/decrease in orbiter processing flow time associated with the proposed modification, in days. (In this case, x_3 will be positive for a decrease, negative for an increase).

In each case, the positive sign was chosen for what would be expected from a modification. That is, one would expect a processing modification to result in a cost savings for each flow, an initial cost, and a decrease (or at least no increase) in flow time.

Typically, one would expect a processing modification to result in some type of change to existing procedures. Therefore, important variables to consider would be:

NOCCPF = the number of occurrences of the change per flow (e.g., the number of times doors were not required to be removed per flow).

DLHOURSOC = the increase/decrease in the number of direct labor hours per . . occurrence.

MCOSTOC = the increase/decrease in the material cost per occurrence.

FLOWS = the expected number of flows per year.

RATIO = the number of indirect hours required per direct hour of labor.

DLCOST = the cost per hour for direct labor.

ILCOST = the cost per hour for indirect labor.

Note that several of these variables would typically be uncertain quantities, especially NOCCPF, DLHOURSOC, MCOSTOC, and FLOWS.

A Demos model could be employed to compute the expected utility for a proposed modification. The Demos diagram for such a model is shown in Figure 8. In this diagram, the variables, not already defined, are given as:

DLSAVINGS = expected yearly savings in dollars from reduction in direct labor hours.

= NOCCPF * DLHOURSOC * FLOWS * DLCOST.

ILSAVINGS = expected yearly savings in dollars from reduction in indirect labor hours.

= NOCCPF * DLHOURSOC * FLOWS * RATIO * ILCOST

MCSAVINGS = expected yearly material cost savings in dollars.

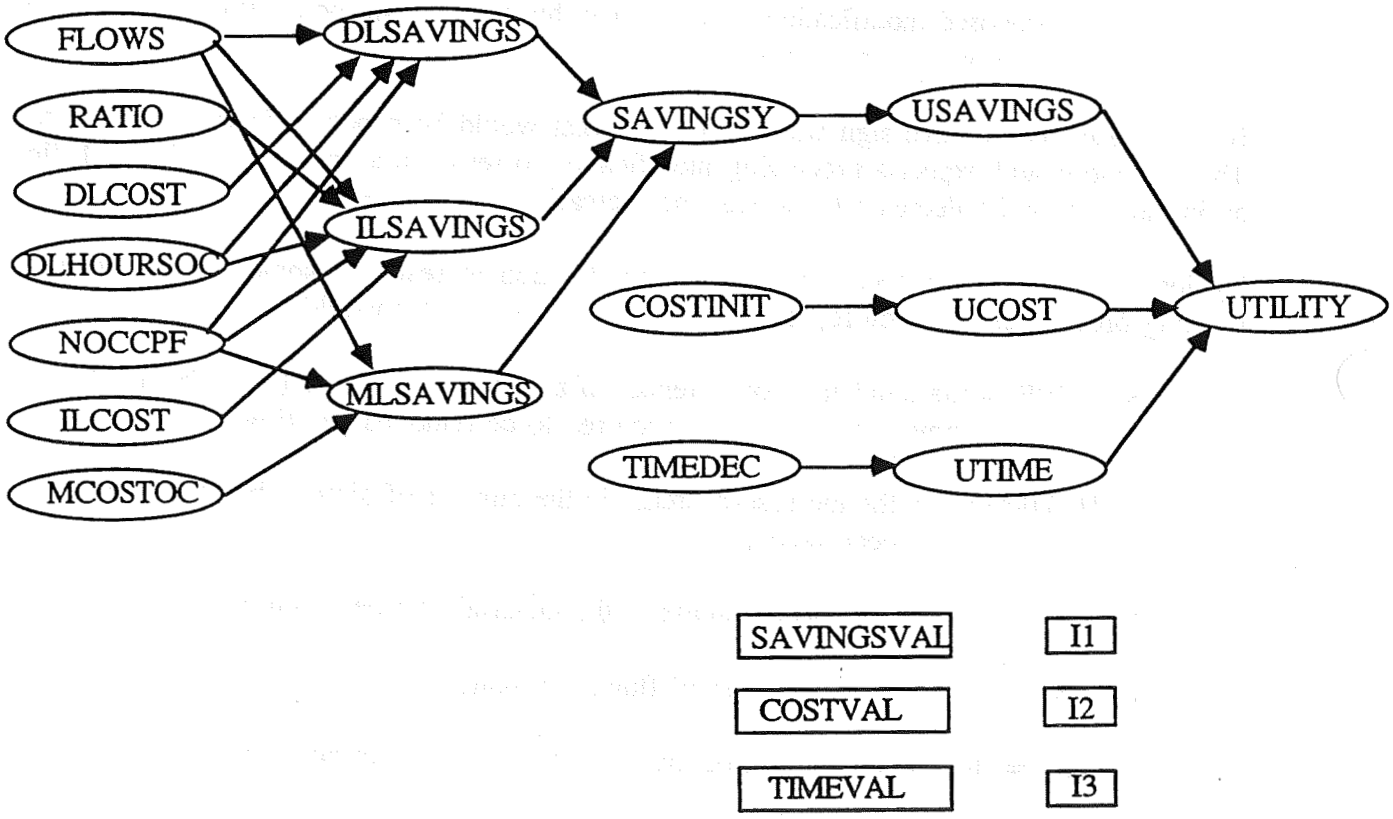


Figure 8. Demos Diagram for Assessment of Orbiter Processing Modifications.

= MCOSTOC * NOCCPF * FLOWS

SAVINGSY = expected yearly savings, in dollars.

= DLSAVINGS + ILSAVINGS + MCSAVINGS

COSTINIT = initial expected cost associated with the modification.

TIMEDEC = expected decrease in orbiter flow time, in days as a result of the modification.

Note that the individual attribute utility function values are computed at the nodes with labels USAVINGS, UCOST, and UTIME for the variables x_1, x_2, and x_3, respectively. The functional values themselves are computed through linear interpolation. The points on the individual attribute utility function graphs are input as values through the nodes labeled SAVINGSVAL and I1 for $u_1(x_1)$, COSTVAL and I2 for $u_2(x_2)$, and TIMEVAL and I3 for $u_3(x_3)$.

Finally, the overall utility function value is computed at the node labeled UTILITY, and is itself a function of the three individual attribute utility functions. A multiplicative form for the utility function was used, as follows:

$$\begin{aligned} u(x_1, x_2, x_3) = {} & k_1u_1(x_1) + k_2u_2(x_2) + k_3u_3(x_3) \\ & + k_{12}u_1(x_1)u_2(x_2) + k_{13}u_1(x_1)u_3(x_3) \\ & + k_{23}u_2(x_2)u_3(x_3) \\ & + k_{123}u_1(x_1)u_2(x_2)u_3(x_3) \end{aligned}$$

Note that this form can be easily reduced to the additive form by settling all of the scaling constants except for k_1, k_2, and k_3 equal to 0.

Choosing Between the NC Machining and The Splash Manufacturing Processes for Orbiter Tiles

After each flight of an orbiter approximately 20 to 100 of the orbiter tiles must be replaced. Two different processes are available for the manufacture of tiles: an NC machining process and a splash process. Individual TPS engineers make the decision as to which process to choose. The decision is currently made based upon the individual engineer's intuition about the cost and time required to manufacture the tile by each type of process.

It was thought by the IE Group at KSC that employing some of the techniques of decision analysis would be helpful to the decision making process. The first step was to understand the process as it occurs now. This was accomplished through interviews with various TPS engineers and through the gathering of data associated with past studies by the IE Group. The results of this data gathering process are illustrated in Figures 9 and 10.

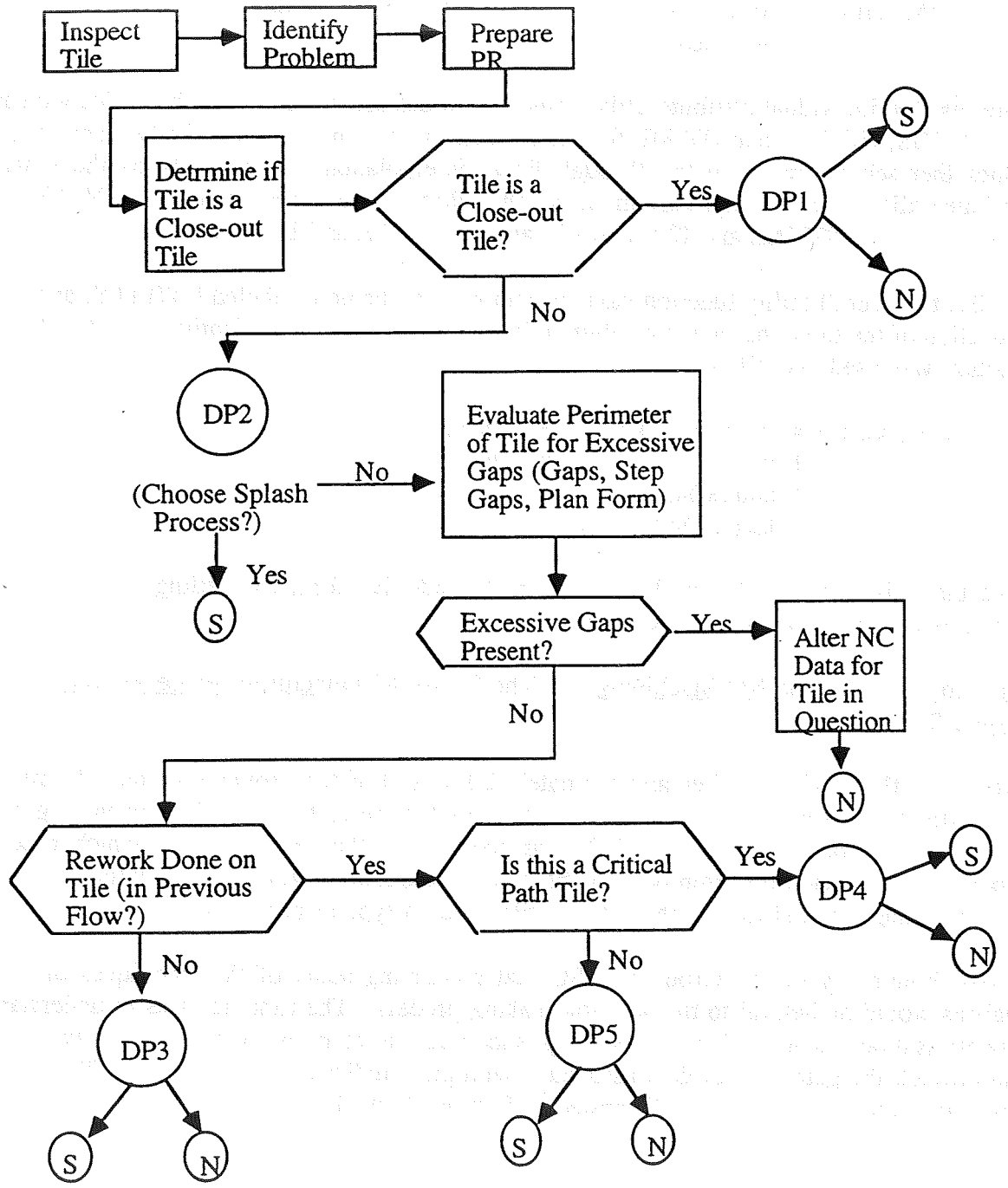


Figure 9. Overall Decision Process for Splash vs. NC Tile Manufacturing Decision.

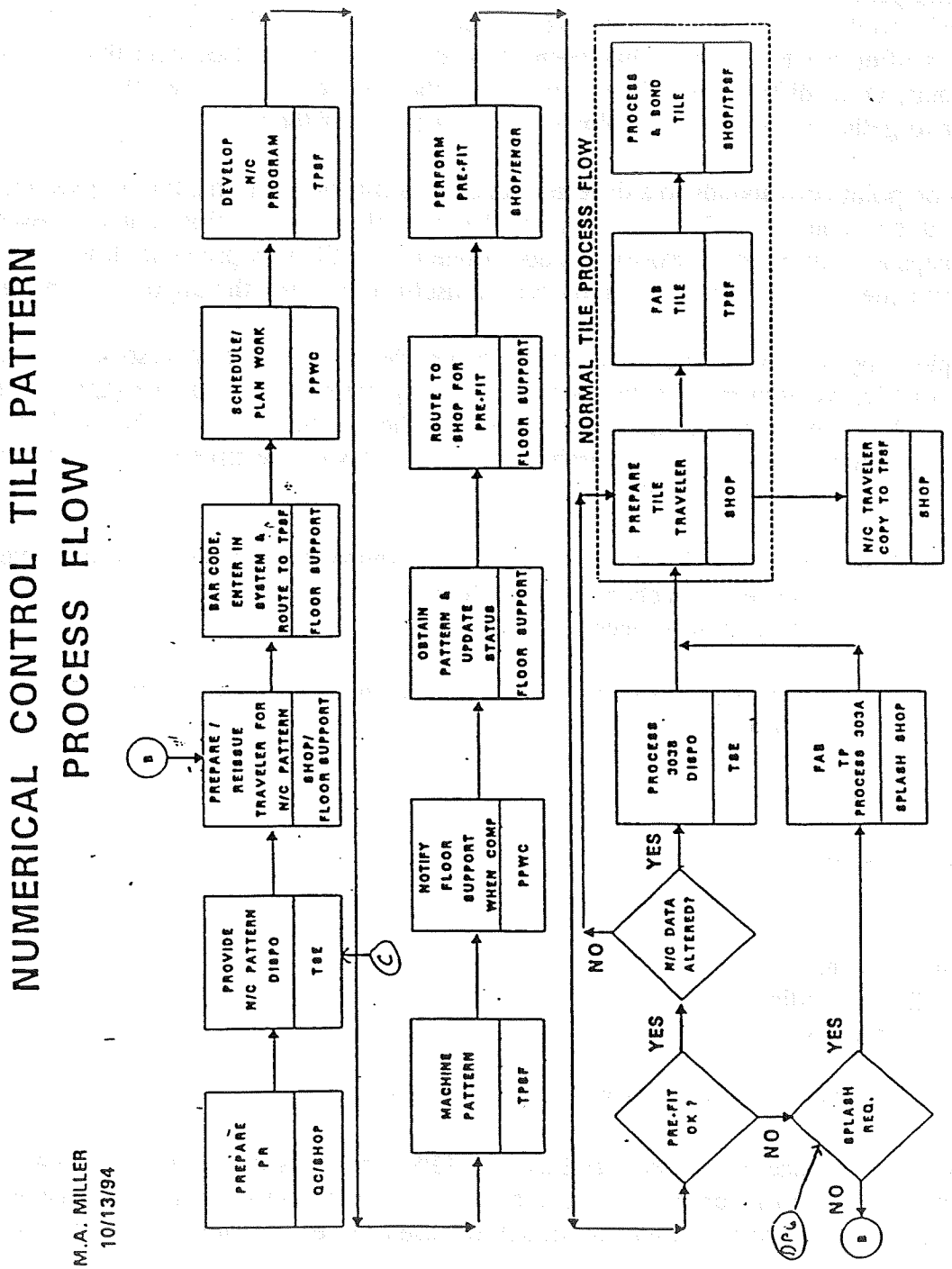


Figure 10. Numerical Control Tile Pattern Process Flow.

The decision as to which manufacturing process to choose can be made at six different points in the overall process, numbered DP1, DP2, ..., DP6 in the relevant Figures. Note that "S" and "N" refer to the splash and NC process, repectively, and note that the last decision point, DP6, occurs after the NC process has been tried at least once. Also, note that the second decision point, DP2, differs from the others in that the decision is to either choose the splash process, or to gather more data by evaluating the perimeter of the tiles.

Each decision point corresponds to a different set of information about the tile in question. For examples, at decision point DP4, one knows that the tile is not a close-out tile, excessive perimeter gaps are not present, rework has been done on the tile in a previous flow, and this is a critical path tile. All of this information can be useful in making the appropriate decision.

For example, by gathering and analyzing data on the time and cost associated with the production of tiles by each manufacturing process, categorized by type of tile (hard, medium, and easy) and by point in the overall process at which the decision is made (DP1 to DP6), one can develop evaluative models giving probabilistic estimates of the time and cost required to manufacture a tile. More specifically, let

$f_t(x,y,z)$ = a density function over the time required to manufacture a tile given that process x is chosen, for a tile of type y, for a decision made at point z in the overall process.

$f_c(x,y,z)$ = a density function over the cost required to manufacture a tile, for x, y, and z.

In the above functions

x = 1, for the splash process
    2, for the NC process

y = 1, for hard
    2, for medium
    3, for easy

z = i, for decision point DPi for i = 1, 2, ..., 6.

In an ideal situation, one could assess and use the TPS engineer's utility function to allow him to choose the manufacturing process that would maximize his expected utility over time and cost. The expected utility would be computed through the use of three functions: the utility funciton, f_t, and f_c.

A more likely approach, given that the appropriate data could be collected, would be for the TPS engineer to just use the f_t and f_c functions as inputs to his decision process. This input could work in the following fashion.

Suppose that the TPS engineer is at decision point DP4 in the decision process, for a "medium" tile. This means that the tile in question is a critical path tile, rework was done or the tile in a previous flow, excessive perimeter gaps are not present, and the tile is not a close-out title. The data analysis has indicated that tiles of this type require, on average, 20 days to manufacture with a standard deviation of 4 days using the splash process, and 18 days to manufacture with a standard deviation of 6 days using the NC process. The TPS engineer would then choose one of the two processes based upon his time preferences and the amount of risk he would be willing to take.

Note that instead of presenting the information concerning manufacturing time in terms of mean and standard deviation, this information could also be presented in terms of probability bands (e.g., 10% probability of taking longer than 24 days, etc.).

Another approach for modeling this decision process would be to use a decision tree (French, 1986), since what's involved in this situation is a sequence of decisions to be made over time.

At the time of the writing of this report, the appropriate data to perform the types of analyses described here was still in the process of being collected.

7. BRIEF SUMMARY

The techniques of decision analysis and related methodologies can be important tools for the Industrial Engineering Staff at the Kennedy Space Center. This report has described how well-structured problems can be constructed from ill-structured problems, and how the various methodologies of decision analysis can be applied in solving well-structured problems. Important problem characteristics which affect the choice of modeling methodologies are also discussed. Finally, specific applications at the Kennedy Space Center are described.

References

VanGundy, A.B., Techniques of Structured Problem Solving, 2nd Edition, Van Nostrand Reinhold Co., New York, 1988.

Miser, H.J. and E.S. Quade (Editors), Handbook of Systems Analysis: Overview of Uses, Applications, and Practice, Elsevier Publishing Co., New York, 1985.

Ackoff, R.L., Redesigning the Future: A Systems Approach to Societal Problems, John Wiley and Sons, New York, 1974.

Basadur, M., S. Ellspermann, and G.W. Evans, "A New Methodology for Formulating Ill-Structured Problems," Omega 22, 627-645, 1994.

Wilson, G., Problem Solving and Decision Making, Kogan Page Limited, London, 1993.

Schoennauer, A.W.W., Problem Finding and Problem Solving, Nelson-Hall, Chicago, 1981.

French, S., Decision Theory: An Introduction to the Mathematics of Rationality, John Wiley and Sons, New York, 1986.

Keeney, R.L. and H. Raiffa, Decisions with Multiple Objectives: Preferences and Value Tradeoffs, Second Edition, Cambridge University Press, New York, 1993.

Pegden, C.D., R.E. Shannon, and R.P. Sadowski, Introduction to Simulation Using SIMAN, McGraw-Hill, Inc., New York, 1990.

Law, A.M. and W.D. Kelton, Simulation Modeling and Analysis, Second Edition, McGraw-Hill, New York, 1991.

Pritsker, A.A.B., Introduction to Simulation and SLAM II, Fourth Edition, John Wiley, New York, 1995.

Lumina Decision Systems, Inc., Demos User's Reference, published by Lumina Decision Systems, Inc., 1993.

Linstone, H.A. and M. Turoff (Editors), The Delphi Method: Techniques and Applications, Addison-Wesley, Reading, Mass., 1975.

Shachter, R.D., "Evaluating Influence Diagrams," Operations Research 34, 871-882, 1986.

Bunn, D.W., Applied Decision Analysis, McGraw Hill, New York, 1984.

Saaty, T.L., Decision Making for Leaders: The Analytical Hierarchy Process for Decisions in a Complex World, Second Edition, University of Pittsburgh, Pittsburgh, PA 1988.

Nemhauser, G.L., Optimization, North Holland, New York, 1989.

Evans, G.W., "Using Demos as a Modeling Tool for Decision Problems," NASA/ASEE Technical Report, Processing and Enhancements Division, Kennedy Space Center, June 1995.

1995 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

S7-47
7747
p. 30

LIGHTNING FORECASTING STUDIES USING LDAR, LLP, FIELD MILL, SURFACE MESONET, AND DOPPLER RADAR DATA

Dr. Gregory S. Forbes, Associate Professor
Mr. Steven G. Hoffert, Graduate Student
Department of Meteorology
The Pennsylvania State University
University Park, Pennsylvania

KSC Colleague - Carl Lennon
Atmospheric Science

Contract Number NASA-NGT-60002
Supplement 19

August 8, 1995

ACKNOWLEDGEMENTS

Participation in the NASA/ASEE Summer Faculty Fellowship Program has been an enjoyable and profesionally stimulating experience. Appreciation is extended to all those who support this program financially, and to the many individuals who made presentations or gave tours to the Faculty Fellows. Particular thanks are extended to Dr. Ray Hosler, Co-Director of the NASA/ASEE Summer Faculty Fellowship Program, and to Kari L. Stiles, Administrative Assistant, who coordinated the program extremely well.

It was a great pleasure to work with Carl Lennon and others associated with the Lightning Detection and Ranging (LDAR) system in TE-ISD. Carl Lennon is to be congratulated for the development of the LDAR system, which shows great capabilities in revealing the presence of electrical discharges within, and emanating from, thunderstorms. His support and encouragement of the applied research reported here is appreciated. The authors benefitted from discussions with Launa Maier, who has vast knowledge of thunderstorms and their electrical processes.

The authors are very appreciative of the efforts of the NYMA employees, Steve Schaefer and Phil Gardner, who operate the LDAR system and archive its data, and patiently dealt with many data requests and the increased computer load while the data were being processed. They, together with TE-ISD, made computer resources available for LDAR data processing. Steve was particularly helpful in retrieving data for the studies described below.

Many groups and individuals helped provide companion data sets for use in the study. Bart Hagemeyer (MIC) and Dave Sharp (SOO) all of the National Weather Service Office in Melbourne, assisted in WSR-88D radar data acquisition. Phil Gemmer of EG&G and Hal Herring, Susan DeRussey and Val Valek of CSR provided field mill and wind tower (surface mesonet) data. Gerry Talley of TE-COM-2 provided LLP data. John Manobianco and Mark Wheeler of the Applied Meteorology Unit (AMU; ENSCO) are helping acquire wind tower data.

Graduate student Steven G. Hoffert of Penn State participated in the summer research project. He took the lead in radar data retrieval and overlaying LDAR and other data on the radar imagery and was liason with the National Weather Service Office in Melbourne in their process of learning to use LDAR data in operational weather forecasting. Over the past year he and Matt Pearce, another Penn State graduate student, developed the radar and overlay display software used in the research. John Madura of the Weather Projects Office (TM-SPO-3) provided financial support for the participation of Steve Hoffert.

The processing of large quantities of data from diverse sources required considerable computer support. Phil Gemmer, Lloyd Albright, Mark Sorger, Carol Lundberg, and Cheryl Axelson of EG&G provided exceptional service in loading data onto the KSC computer system (KSDN). Carl Lennon of TE-ISD made a workstation available to Steve Hoffert. John Kiriazes and Steve Tam of TE-COM helped provide access to the Internet, and helped with other computer problems. John Diercks and Art Person of Penn State provided computer assistance via phone and e-mail. Workstations in the Wave Analysis Lab were used to display radar data and LDAR overlays, and the assistance of Tom Henning (NGTS) is appreciated.

The authors looks forward to a continued interaction with Carl Lennon, Launa Maier, Steve Schaefer, Phil Gardner, and others using LDAR data for weather applications, including Robin Schumann (ENSCO) and Frank Merceret of the AMU/TM-SPO. Our goal is to help the operational weather forecasters of the 45th Weather Squadron--detached out of Patrick AFB, USAF Space Command--who provide KSC with operational weather support from the Cape Canaveral Forecast Facility.

ABSTRACT

The ultimate goal of this research is to develop rules, algorithms, display software, and training materials that can be used by the operational forecasters who issue weather advisories for daily ground operations and launches by NASA and the United States Air Force to improve real-time forecasts of lightning. Doppler radar, Lightning Detection and Ranging (LDAR), Lightning Location and Protection (LLP), field mill (Launch Pad Lightning Warning System--LPLWS), wind tower (surface mesonet) and additional data sets have been utilized in 10 case studies of thunderstorms in the vicinity of KSC during the summers of 1994 and 1995. These case studies reveal many intriguing aspects of cloud-to-ground, cloud-to-cloud, in-cloud, and cloud-to-air lightning discharges in relation to radar thunderstorm structure and evolution. They also enable the formulation of some preliminary working rules of potential use in the forecasting of initial and final ground strike threat. In addition, LDAR and LLP data sets from 1993 have been used to quantify the lightning threat relative to the center and edges of LDAR discharge patterns.

Software has been written to overlay and display the various data sets as color imagery. However, human intervention is required to configure the data sets for proper intercomparison. Future efforts will involve additional software development to automate the data set intercomparisons, to display multiple overlay combinations in a windows format, and to allow for animation of the imagery. The software package will then be used as a tool to examine more fully the current cases and to explore additional cases in a timely manner. This will enable the formulation of more general and reliable forecasting guidelines and rules.

SUMMARY

A diversity of data types have been used to seek to augment the meteorological bases upon which lightning forecasts are issued. The goal of this research is to develop rules, algorithms, display software, and training materials that can be used to improve the real-time lightning advisories issued for daily ground operations and launches by NASA and the United States Air Force (USAF).

Research during 1995 built upon studies conducted during the previous two years. Over that period we sequentially developed capabilities to use Lightning Detection and Ranging (LDAR) data in conjunction with Lightning Location and Protection (LLP), Doppler radar, sounding, wind tower (surface mesonet), wind profiler, and field mill (Launch Pad Lightning Warning System--LPLWS) data. Beginning in summer, 1994, we developed computer software to overlay these data sets and display them as color images. In summer, 1995, we used these tools to pursue a number of case studies aimed at providing insight toward the solution of several lightning forecasting challenges.

One of the key lightning forecasting challenges is to anticipate where and when lightning will appear in the vicinity of the NASA and USAF warning sites. Surface mesonet data can provide useful information by highlighting regions of convergence of the wind, where dynamic forcing can augment the buoyant process of thunderstorm formation. This type of forcing is most effective when the convergent location is colocated with a local temperature (buoyancy) maximum. The convergence can be a result of the sea breeze front, the outflow from some previous or nearby convective storm, or other phenomena. We have found that mappings of the temperature and wind vector change over short (15 to 30-minute) time intervals can highlight areas where local hot spots are forming and where the air is becoming more convergent. Radar and satellite can also show the location where clouds are growing largest and tallest. We have found that once the radar reflectivity of the storms reaches 30 dBZ at or above the 8 km level, then the development of lightning can be expected if the storm remains active. The existence of storm-top divergence using Doppler radar signifies that the storm continues to have updraft and should proceed to generate lightning.

Another key lightning forecasting challenge is to determine when the threat of cloud-to-ground lightning has ended. In many situations anvil and other cloud debris--generated within convective towers and then either carried downwind of slow-moving storms or left behind by storms moving faster than the winds aloft--remains electrically active for tens of minutes to hours over points where the ground strike threat has not yet begun or has long ago ended, respectively. These clouds can be affiliated with large surface electric fields and with numerous discharges aloft detected by LDAR. This research seeks ways to identify where and when the threat of ground strikes is sufficiently small that phase 2 lightning advisories are not needed. Our studies to date indicate that ground strikes are rare beyond regions where the 20 dBZ reflectivity contour reaches the ground, even if there is higher reflectivity aloft. However, there are notable exceptions. Also, lightning activity may cease once no portion of the storm or adjacent storm has reflectivity of more than 30 dBZ above about 7 km. Adjacent storms must be monitored because cloud-to-cloud discharges initiated in the elevated regions of active storms apparently travel along the anvil or other cloud layers to impact weaker cells nearby.

A third key lightning forecasting challenge is to know where and when ground strikes occur within active convective storms. How far beyond the edge of the current LDAR discharge pattern, or radar echo, or cloud does there remain a non-negligible threat of an imminent ground strike? This topic is also addressed. Probabilities of ground strikes per square kilometer per minute are computed as a function of radius from the center of the composite LDAR discharge pattern. In most cases, ground strikes do not occur more than about 2 km beyond the edge of where the 20 dBZ reflectivity contour lies in the lowest 4 km.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

I. INTRODUCTION

The Kennedy Space Center (KSC) is located in one of the regions of the United States (and even the world) that encounters the most lightning strikes to ground per unit area (refs. 1,2,3). The possibility of lightning at the surface or aloft is, of course, a hazard that must be avoided during launches. On a daily basis, however, there are many operations at KSC which must be curtailed if there is a threat of a lightning strike to ground in the vicinity. The accuracy and timeliness of lightning advisories, therefore, has both safety and economic implications. The ultimate goal of the research described in this report is to provide information that can be used to improve the process of real-time detection and warning of lightning by weather forecasters who issue lightning advisories.

1.1 DATA SETS USED IN THE RESEARCH

Special networks of remote sensing equipment have been established to provide highly accurate information concerning lightning in the vicinity of KSC: the Lightning Location and Protection (LLP) system, the Lightning Detection and Ranging (LDAR) system, and a Launch Pad Lightning Warning System (LPLWS). In addition, a Catenary Wire Lightning Instrumentation System (CWLIS) detects electrical surges in wires at the launch pads when struck by lightning. The first two systems detect lightning signatures. LPLWS, by contrast, responds not only to lightning but also detects electric fields at the surface induced by electrified clouds, thunderstorms, and other atmospheric conditions. Data from the LLP, LPLWS, and LDAR systems were used in this study.

The LLP system (ref 4) detects lightning ground strikes through use of a network of magnetic direction finding antennae which sense electromagnetic disturbances triggered by lightning in a broad band of frequencies. Individual antennae detect a particular ground strike at different azimuth angles, and the location of the ground strike is essentially determined by finding the point of intersection of lines drawn from the antennae toward the source of the disturbance. The LLP system is approximately 90% efficient in detecting ground strikes near KSC, with position accuracy of about 1 km.

The LDAR system was developed by Carl Lennon and colleagues at KSC TE-CID-3 (ref 5). Its antennae detect lightning-induced disturbances at 66 MHz (VHF) frequency of the stepped-leader type. This system uses a time of arrival (TOA) approach, and achieves extremely accurate timing through use of the Global Positioning System (GPS). The lightning-induced disturbance, travelling at the speed of electromagnetic propagation, arrives at different antennae at slightly different times. The three-dimensional position of the lightning source is determined by essentially converting these time offsets into distance differences, and then performing a triangulation. The LDAR system began real-time operation in June, 1992.

The LDAR system can generate up to 10,000 data points per second, yielding numerous data points per lightning flash. Tests of the position accuracy of the LDAR data by Launa Maier have shown that within 10 km of the central antenna, 95% of the data points are accurate to better than 200m, and 50% are accurate to better than 100m.

Companion data sets included Doppler radar data from the National Weather Service WSR-88D at Melbourne, Florida, wind tower (surface mesonet) data, and rawinsonde data from Cape Canaveral. Future research will include the use of data from the National Center for Atmospheric Research CP-2 multiparameter radar that operated from 3 July through 17 August 1995. This dual-wavelength, dual-polarization radar allows information to be gleaned regarding the form of storm precipitation (ice crystal, small or large liquid droplet, hail). Its scan strategy was

normally tailored to focus on storms over the wind tower and field mill networks and to adequately sample the entire storm volume. Unfortunately, as will be seen in examples below, the WSR-88D three-dimensional storm coverage is often not totally adequate to accurately depict the upper portions of storms. (This is because the time-consuming need to monitor a full 360° of azimuth only allows a limited number of elevation angles to be scanned.)

1.2 PAST RESULTS AND CURRENT OBJECTIVES

This was the third summer of study involving research to utilize Lightning Detection and Ranging (LDAR) data, together with companion data sets, aimed at developing rules, algorithms, and training materials that can be used by the operational weather forecasters who issue weather advisories for daily ground operations and launches by NASA and the United States Air Force. Research during 1993 (ref 6) enabled the development of a computerized scheme for clustering the LDAR data into groups of data points associated with individual thunderstorms (as described above), tracking these LDAR-defined storms, and comparing the positions of the LDAR-detected lightning to those of other remote sensing systems. It was determined that LDAR-detected discharges aloft within the storm precede ground strikes by about 5 minutes in the region within 60 km of KSC, making LDAR a very useful tool for issuing very-short-term weather advisories and warnings. By recognizing and including storm movement in a forecast scheme, mappings of current LDAR data points can, to some degree of accuracy, be extrapolated with a storm motion vector to make forecasts of future cloud-ground strikes.

Research during 1993 showed, however, that beyond about 10 minutes areal storm growth and the development of new thunderstorm cells (as opposed to the translation of steady-state lightning patterns) became increasingly important factors in the prediction of future lightning ground strikes. Hence, one focus of the 1994 research (ref 7) was to include a storm growth factor in the forecast scheme. The 1994 study showed that only modest gains in forecast accuracy could be achieved through use of a symmetric growth factor, and that the sudden developments or decays of thunderstorm cells and their evolutions ultimately limited the accuracy of forecasts beyond about 10 minutes. Thus, research was begun to explore ways to use companion meteorological data sets and develop forecast schemes to help the forecasters anticipate new thunderstorm formation. In addition, forecasters must determine when the lightning threat at a site has ended--a task made difficult because electrified anvil clouds are often left behind above a site long after the core of the storm has exited the region. Thus, work was begun to examine this problem. Much of the work entailed the development of software to analyze and generate overlay displays of the various data sets.

The objectives of the 1995 research were to use the software that had been previously developed to perform a number of case studies, using 1994 and 1995 data. The goals of the case studies were to (1) examine relationships between lightning and storm structure as depicted by radar, (2) identify signatures in other data sets that can be useful in the forecasting of lightning, and seek guidelines and rules that forecasters might use in determining (3) where and when lightning ground strikes will first occur at a forecast site and (4) when the ground strike threat has ended at a forecast site. Table 1-1 lists the dates and character of the cases examined to date.

1.3 USE OF MULTIPLE DATA SETS IN LIGHTNING FORECASTING

Though each of the data sets mentioned contributes valuable information, none of the data sets alone are adequate to supply answers to all forecast decisions. Thus, the emphasis of this research will be to use multiple data sets in order to come up with the best overall forecast guidance.

TABLE 1-1

CASES EXAMINED USING RADAR, LDAR, LLP, LPLWS

10 June 1994, Day 161 --	Storm initiation on mesoscale boundary; quasi-horizontal discharges
1 July 1994, Day 182 --	Storm formation
18 July 1994, Day 199 --	Storm formation
19 July 1994, Day 200 --	Storm formation; impact of peripheral storms on LPLWS readings
29 July 1994, Day 210 --	Thunderstorm initiation; microburst storm
11 June 1995, Day 162 --	Storm formation through layered cloud at end of storm
12 June 1995, Day 163 --	Storm formation; end of storm cloud layer
20 June 1995, Day 171 --	Mostly lightning aloft
20 July 1995, Day 201 --	Storm formation
21 July 1995, Day 202 --	Storm formation through layered cloud at end of storm

When the forecaster is dealing with the problem of when to issue phase 2 lightning warnings (cloud-ground lightning imminent), the case studies suggest that there are limitations to the stand-alone capabilities of individual systems. (1) LDAR and LLP inherently only show the existence of lightning, but can help a forecaster anticipate when lightning from an approaching storm might become a threat at a particular site. (2) Furthermore, while the first LDAR discharges in a storm are typically aloft, and thus give a 4-5 minute lead time relative to ground strike threat, in about 25% of the storms the first ground strike occurs within a minute of the first appearance of LDAR. (3) In the cases studies, there are a number of instances where LDAR detects cloud-ground discharges that LLP does not detect. According to Launa Maier (personal communication), this tends to occur when the cloud-ground discharge is not very energetic. (4) However, LDAR also can fail to detect some ground strikes because of the dominance of the return stroke discharge at low levels (whereas LDAR detects stepped leader discharges). (5) Field mill readings can be large over a site in some instances long prior to any valid ground strike threat, due to electrified cloud blowoff from upwind storms or due to other low-level meteorological conditions. (6) On other occasions, with rapid storm development or small-diameter storms, the field mill readings may only become large briefly before lightning activity begins. Thus, the use of radar and surface mesonet data can yield invaluable clues regarding the location of thunderstorms and the state of their progress toward electrificaiton.

When the forecaster is dealing with the problem of when to terminate the phase 2 lightning warnings (imminent ground strike threat has passed), case studies again indicate that no individual data set is adequate. (1) LDAR often shows the existence of extensive quasi-horizontal discharge patterns aloft long after ground strikes have ceased. (2) LLP may misrepresent some discharges aloft as ground strikes. (3) LDAR may fail to detect some ground strikes, as noted above. (4) Field mill readings can remain large long after the ground strike threat has ceased, due to lingering electrified cloud layers overhead or nearby.

II. GROUND STRIKE PROBABILITY IN RELATION TO LDAR PATTERN AND RADAR ECHO

Results from the 1993 studies (ref 6) indicate that within the concurrent minute, 85% of the LLP-detected ground strikes occurred within the bounds of the LDAR discharge pattern, and 98% of the ground strikes occurred within or no more than 2 km beyond the edge of the LDAR pattern. These data were combined with storm diameter to come up with a composite risk analysis for ground strike threat during the next minute relative to the current LDAR pattern. Ground strike probability has been computed as a function of radius from the LDAR pattern center per unit area within a series of concentric rings, and is plotted as Figure 1-1. Also shown is the edge of the composite LDAR data pattern. The absolute probability of a ground strike decreases from about 0.01 $km^{-2}min^{-1}$ near the center of the LDAR pattern, to about 0.001 near the edge of the LDAR pattern, to about 10^{-4} at about 3 km beyond the edge of the LDAR pattern and to about 10^{-5} $km^{-2}min^{-1}$ at about 6 km beyond the edge of the LDAR pattern. Statistics were compiled from 319 LDAR-detected thunderstorms that occurred during 33 hours on 13 days in June and July 1993 and within +/- 52 km west/east and +/- 40 km north/south of KSC.

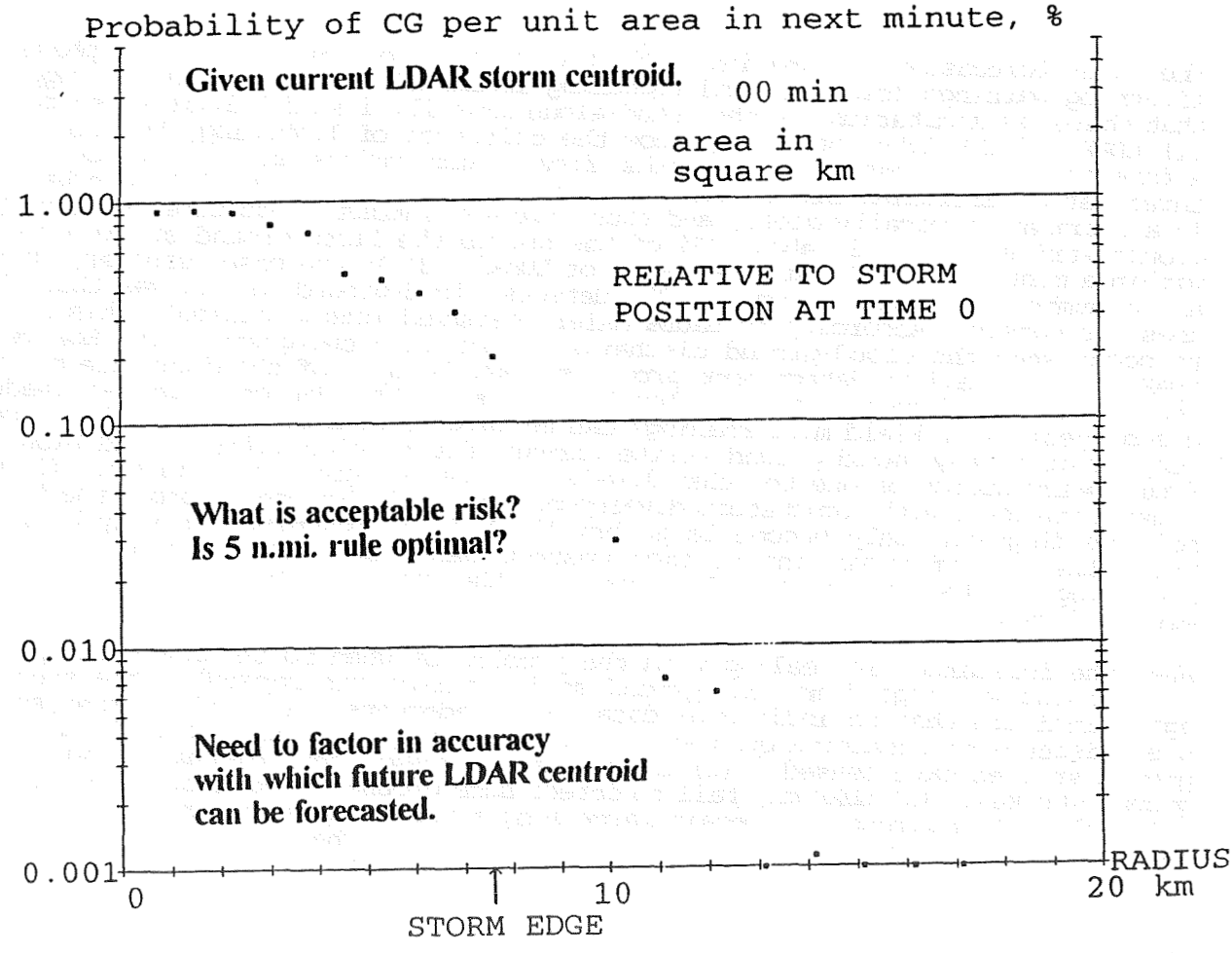


Figure 2-1. Probability of a lightning ground strike within one minute, given the current position of the center of the LDAR discharge pattern, as a function of distance. Probability (y-axis) is in percent per square kilometer. Position of the edge of the composite LDAR pattern edge is indicated by arrow.

During the inspection of the cases to date (Table 1-1) it became obvious that the great majority of ground strikes occur near or within regions experiencing low-level precipitation. As a working rule, lightning rarely strikes ground more than 2 km beyond the edge of the 20 dBZ reflectivity contour in the lowest 4 km. However, exceptions do occur. Figure 2-1 shows the two most extreme cloud-ground strikes identified within the cases examined to date. In these cases the discharge appears to have travelled toward ground in locations outside cloud edge for at least a portion of the path. The ground strike position in these cases is about 6 and 5 km, respectively, from the 20 dBZ position.

It also became obvious that LDAR and LLP do not always give exactly the same numbers and locations of ground strikes. There are daily examples where LDAR shows strings of data points progressing to within a kilometer of the surface, and which are inevitably ground strikes, that are not shown by LLP. Presumably these flashes had weak return strokes that were missed by LLP. Similarly, LLP daily records ground strikes not accompanied by LDAR points near the surface. This is not totally surprising, since LDAR detects stepped leaders and, in contrast, return strokes dominate the low-level portion of ground strikes. However, on some occasions the LLP points may be erroneous representations of discharges aloft as ground strikes.

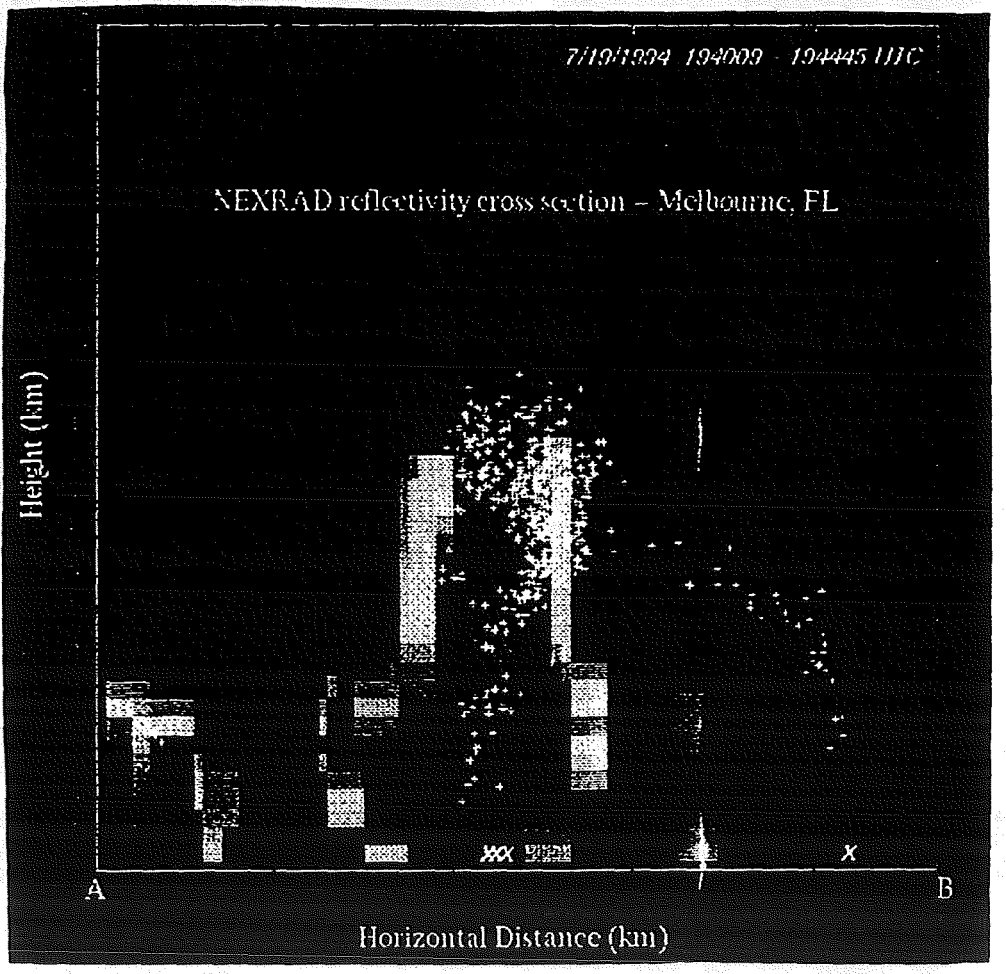


Figure 2-2. Examples of ground strikes in low-reflectivity regions, depicted by LDAR (+ symbols) and LLP (X) overlaid onto cross-section of radar reflectivity. Because the original is in color, reflectivities are hard to distinguish in black and white. Because the radar scanned only at a few elevation angles, storm top in (a) is not depicted.
a. from 19 July 1994

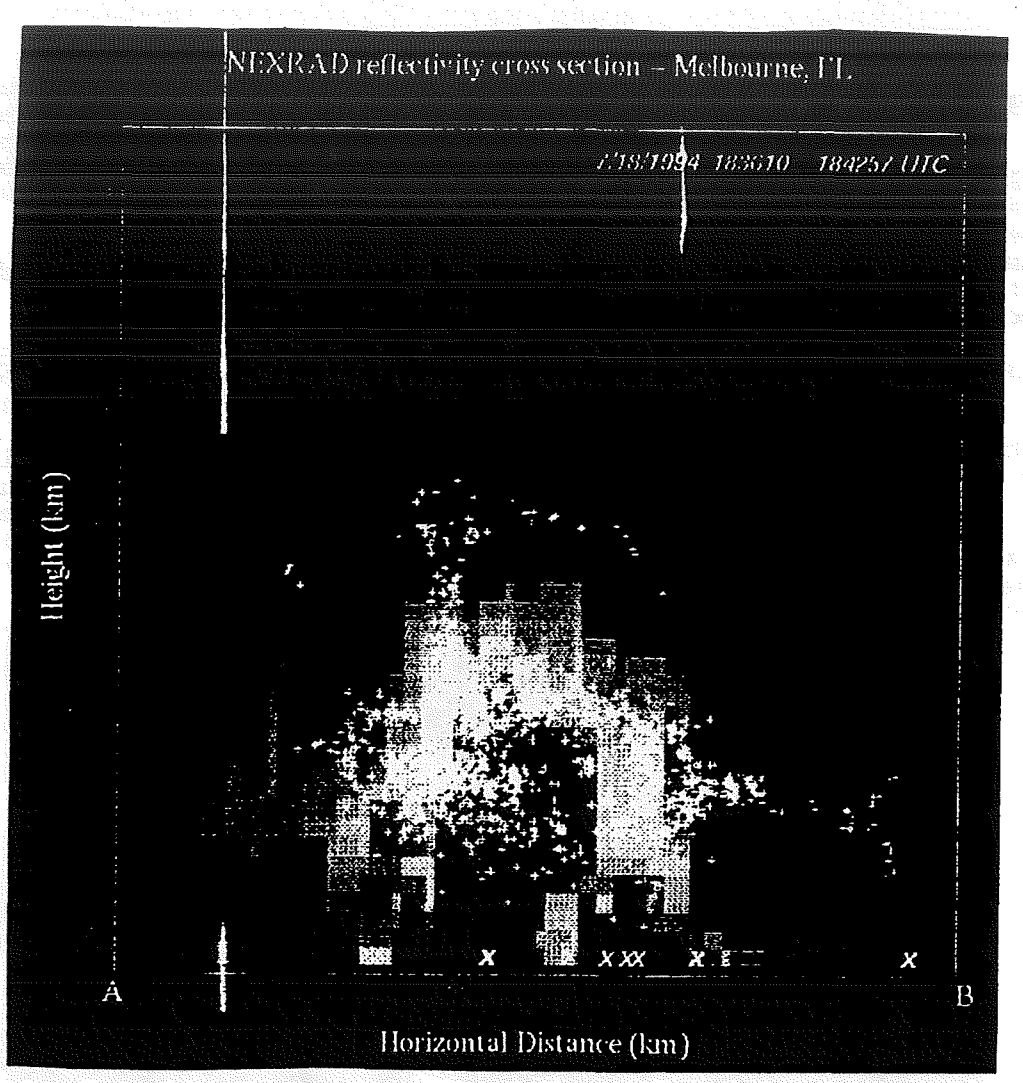


Figure 2-2. Examples of ground strikes in low-reflectivity regions, depicted by LDAR (+ symbols) and LLP (X) overlaid onto cross-section of radar reflectivity. Because the original is in color, reflectivities are hard to distinguish in black and white. Because the radar scanned only at a few elevation angles, storm top in (a) is not depicted.
b. 18 July 1994 .

In order to identify the first and last times of ground strikes affecting (i.e., within 5 n.mi. of) any warning site in the cases studied to date (Table 1-1), the initial and final LLPs of each case were compared to the LDAR data at the time. Many additional discharges were examined by watching the real-time LDAR and LLP displays. Based upon the combined information, an approximate scheme was formulated to identify when an LDAR flash was likely to have yielded a ground strike:

a. There must be at least two LDAR points below altitude of 3 km;
b. At least one of the LDAR points must be below 2.1 km;
c. The lowest LDAR point must be within a slant distance of not more than 2 km from another LDAR point;
d. LDAR points near x = -1.3, y = -1.6 km and below 600 m within the first 0.006 second of each minute must be viewed suspiciously, as these are likely to be due to a calibration pulse emitted from an antenna on the top of the NASA Central Instrumentation Facility building.

III. JOINT USE OF METEOROLOGICAL DATA SETS IN FORECASTING INITIAL LIGHTNING THREAT

Many studies (e.g., ref 8) have shown that the location of thunderstorm initiation can often be inferred through careful use of satellite, radar, and surface data. Radar and satellite can reveal the initial development of small clouds in areas of enhanced ascent due to low-level convergence where the winds from small-scale features such as the sea-breeze, river breeze, or thunderstorm outflow collide with the prevailing large-scale wind. Wind and temperature patterns shown by the KSC wind tower mesonetwork can also reveal these convergence zones and local hot spots of enhanced buoyancy.

The surface mesonetwork can reveal preferred areas of thunderstorm development:

a. Strong convergence precedes new thunderstorm development;
b. New thunderstorm development is enhanced where convergence intersects a thermal (buoyant) plume or thermal boundary (mesoscale front);
c. Development is enhanced where convergence intersects the gradient between a region of temperature change over the last 15 to 30 minutes, with warming on at least one side of the gradient;
d. Sites of future maximum convergence are sometimes revealed by charts showing wind vector change over the last 15 to 30 minutes. Convergence of these streamlines highlight regions becoming more convergent with time.

Figures 3-1 and 3-2 illustrate the above rules, using real data from 18 July 1994. Figure 3-1a shows the surface mesonet data from 1650 UTC, and Fig. 3-1b shows the change in temperature in the straddling period ending at 1700 UTC. Figure 3-2 shows the streamlines and convergence at 1650. The data show a well-defined convergence zone over Merritt Island near KSC between the Banana and Indian Rivers that is also warming with time, whereas the adjacent regions have remained constant or cooled. Temperatures just south of the convergence center were already warm and are becoming even more buoyant. This combination of convergence (which will force upward motion and help trigger cloud formation) and buoyancy suggests a preferred location for thunderstorm initiation over or just west of KSC on this occasion. Vector changes of the wind (not) shown, also indicate a need to focus on this location.

Figure 3-3 shows the first radar composite (from 1704-1708 UTC) with precipitation in the developing thunderstorm. In the preceding 5 minutes the reflectivity has increased from 10 dBZ and is now in excess of 30 dBZ. Though the radar scan strategy is chopping off the top of the echo in the cross-section, reflectivities exceed 30 dBZ at 8 km, suggesting that lightning is imminent in this storm.

Figure 3-4 shows that the storm also has marked storm-top divergence. There is also convergence at low levels near range of 8 km along the cross section. Hence, the storm should have a vigorous updraft, and precipitation-powered electrification should be well underway.

Figure 3-5 shows the radar composite from 1709-1714 UTC. Maximum reflectivity in the storm is now about 50 dBZ and exceeds 40 dBZ at 8 km altitude. Such a storm should be producing lightning. Indeed, LDAR events commenced at 1709 UTC in the layer between 8.2 and 10.4 km just above the reflectivity core, and have begun to discharge outward as superimposed on the imagery. A newer cell to the north does not yet have high enough reflectivity (30 dBZ) or altitude (8 km) to be producing lightning. Though not shown here, one surface field mill exceeded 1 kV/m at 1711 when data first became available, but this storm is in general too young and too small to have impacted many of the field mills at this stage.

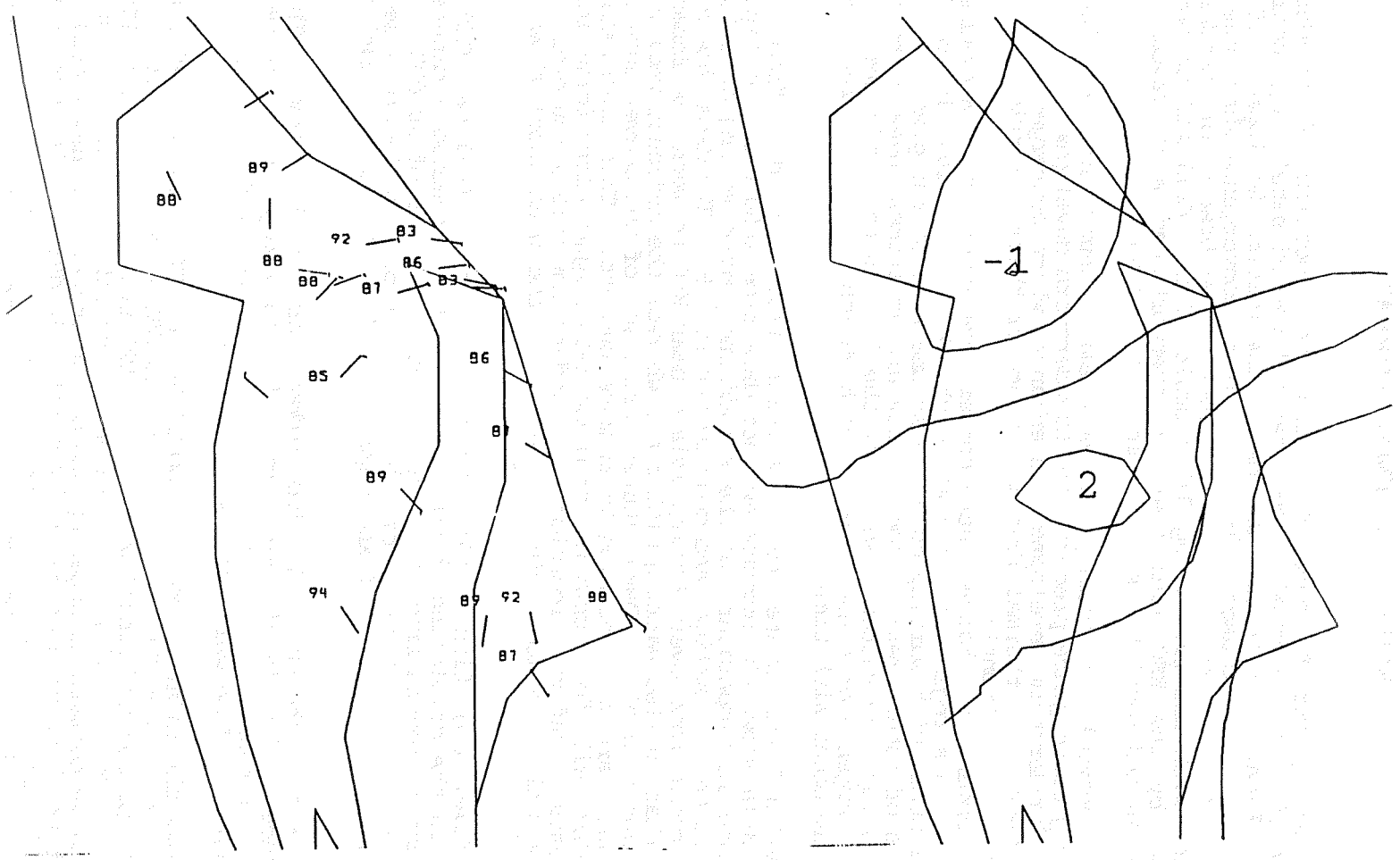
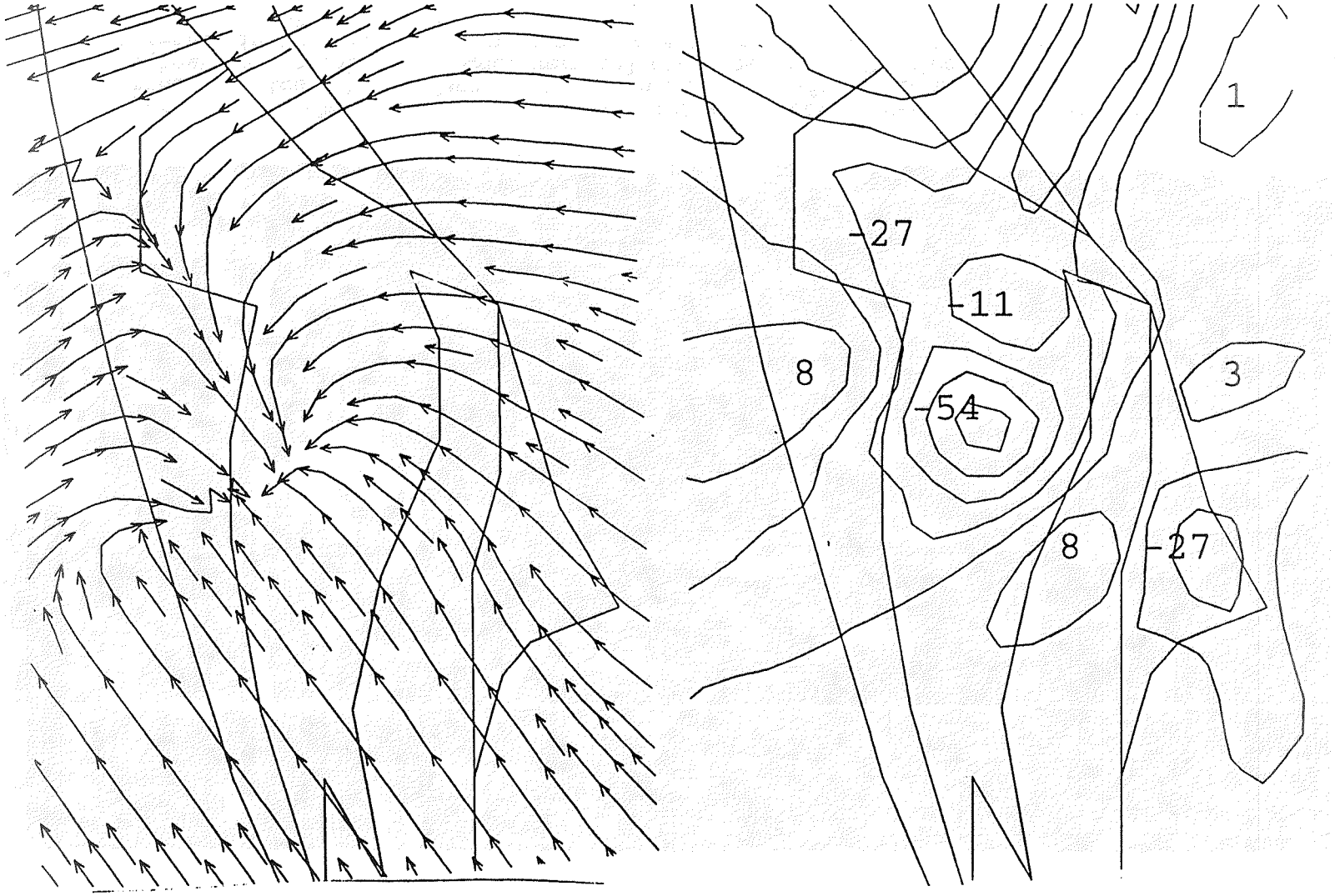


Figure 3-1. Surface mesonet data from 1650 UTC on 18 July 1994 (left) and isotherms of the 15-minute temperature change (right) ending at 1700 (°F).



7/18/94 Day 199 1650 UTC Streamlines and Divergence

Figure 3-2. Streamlines (left) and divergence (right) of the surface wind at 1650 UTC on 18 July 1994. Divergences are in units of 10^{-5} s^{-1}.

Figure 3-3. Radar depiction of the initial storm development, using National Weather Service WSR-88D data from Melbourne, FL. Left, composite reflectivity (highest in the column); right, vertical cross section. Original was in color, so reflectivity shadings do not reproduce well in black and white. Sampled from about 1704-1708 UTC, 18 July 1994.

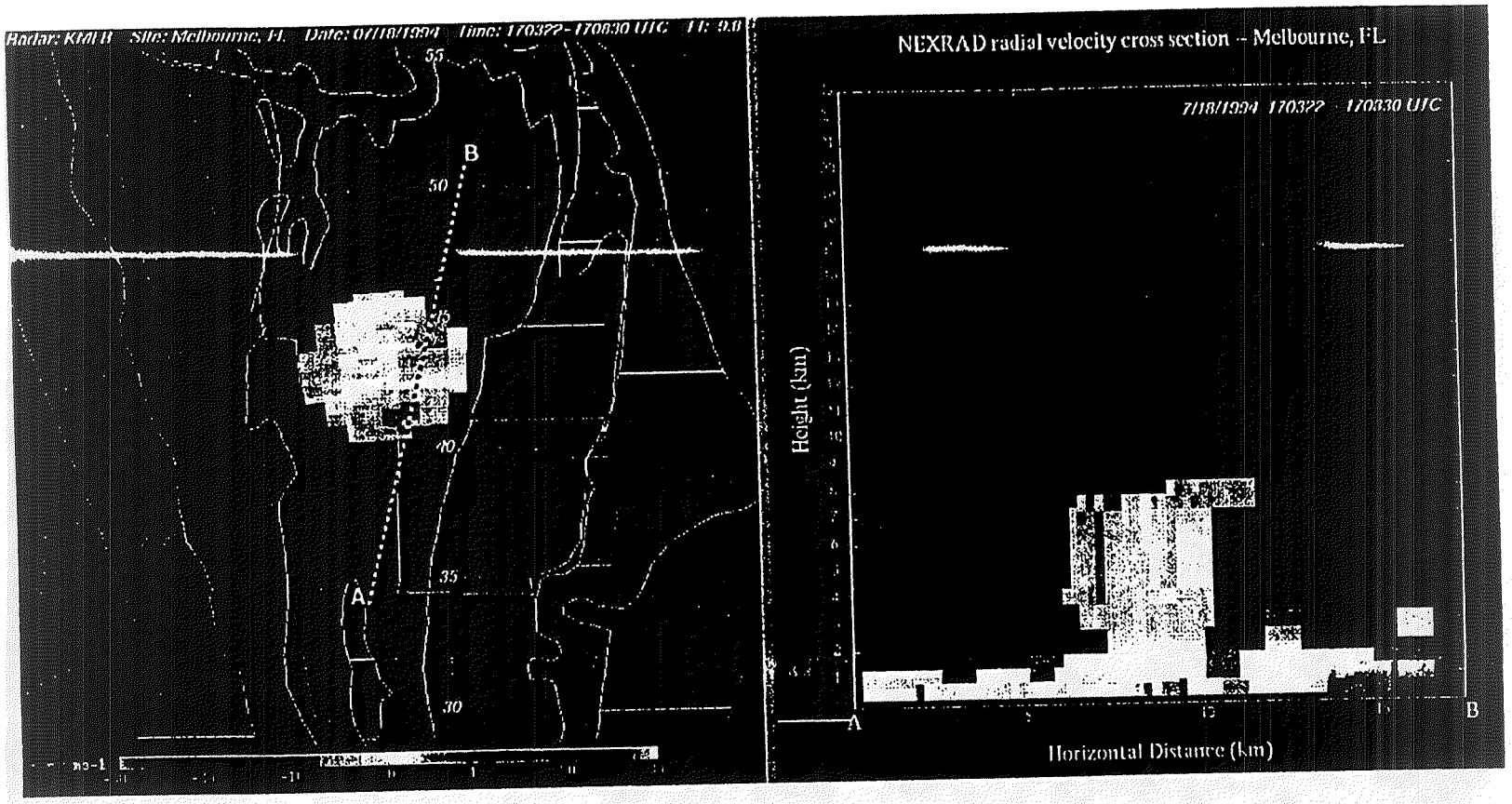


Figure 3-4. Doppler velocities in the storm of Fig. 3-3. Left, velocities in the scan at 9.8° elevation angle are positive (northbound) in the north and negative (southbound) in the southern portion of the storm, indicating divergence. Right, cross-section view.

Figure 3-5. Radar and LDAR depiction of the storm from 1709-1714 UTC. Left, composite reflectivity with LDAR data for this time interval superimposed (white + symbols). Right, cross-section, including LDAR data within 1 km of the cross-section.

Figure 3-6 shows the pair of radar echoes composited for the period 1715 to 1721 UTC, with LDAR data superimposed. The thunderstorm developed ground strikes beginning at 1716, and has also begun to produce quasi-horizontal flashes radiating outward. Figure 3-7 shows that the region of initiation of these quasi-horizontal discharges is within the area of divergence in the upper half of the storm, consistent with being in updraft.

The storm soon began to show a collapsing top, and all LDAR activity ceased at 1743. Case studies suggest that in individual thunderstorms lightning activity begins and ends when the 30 dBZ echo (or greater) appears and then disappears from altitudes of at least 8 km.

IV. JOINT USE OF METEOROLOGICAL DATA SETS IN FORECASTING TERMINATION OF LIGHTNING THREAT

In some ways forecasting when the threat of ground strikes to a region has ended is a more challenging meteorological problem than in predicting where and when the first ground strike will occur. The case studies to date suggest that on many of the former occasions there will be ongoing quasi-horizontal LDAR discharges aloft long after the last ground strike has actually occurred, but the safe assumption is that ground strikes remain a possibility as long as there are LDAR events aloft. Furthermore, the case studies show that surface field mill readings can remain quite high long after the last ground strike has actually occurred, due to the presence of charged anvil or debris cloud overhead or nearby. Unfortunately, there are documented cases in the literature of ground strikes emanating from anvils, so that cautious forecasting is merited. Thus, most rules that would allow the termination of phase 2 advisories in the presence of LDAR discharges or large electric fields implicitly involve accepting some level of risk. Unfortunately, the data base may be inadequate to properly quantify what is the level of risk involved. Certainly we have not yet performed enough case studies to quantify the risk as a function of radar reflectivity, LDAR discharge rate or intensity, or other factors.

Figure 3-8 shows an example of the problem faced by the forecaster. The last ground strike within 5 n.mi. of any warning site in this case was at 2100 UTC. However, there are still LDAR discharges within the clouds aloft, and surface field mill readings remain high. Even by 2130, 16 sites still exceed +\- 1 kV/m. Clearly the site with the 1.7 kV/m field mill reading is not under a very great risk of a ground strike, as there is only a rather thin anvil cloud above it. A working hypothesis has been formulated based upon our case studies, but one that is statistically untested in terms of absolute risk of ground strike: ground strikes are uncommon at points more than 2 km beyond where the edge of the 20 dBZ contour is found within the lowest 4 km of the storm.

Figures 3-9 and 3-10 show another example, where sloping layers of LDAR discharges are found in the reflectivity gradients at the top and bottom of the anvil. The upper LDAR layer of Fig. 3-9 extends farther to the rear of the eastward-moving storm than does the lower layer. Field mill readings at the surface (Fig. 3-10) fall below 1 kV/m beneath the extended LDAR layer. Adoption of the "2 km beyond 20 dBZ" rule above in this case would call for an end to the ground strike threat at about the 24 km range along the cross-section of Fig. 3-9, whereas the upper LDAR layer ends at about the 22 km range.

The cross section of Fig. 3-10 intersects a surface-based slice of reflectivity greater than 20 dBZ, centered near 30 km along the section. A simple adoption of the hypothetical rule above would terminate the ground strike threat at about the 26 km range on the section. However, intercomparison with Fig. 3-9 shows that this cloud layer is not the one being straddled by the layered LDAR points. Indeed, there are no LDAR points along the cross section in this region of enhanced reflectivity. It would be tempting, then, to terminate the ground strike threat at about the 39 km range on the cross section.

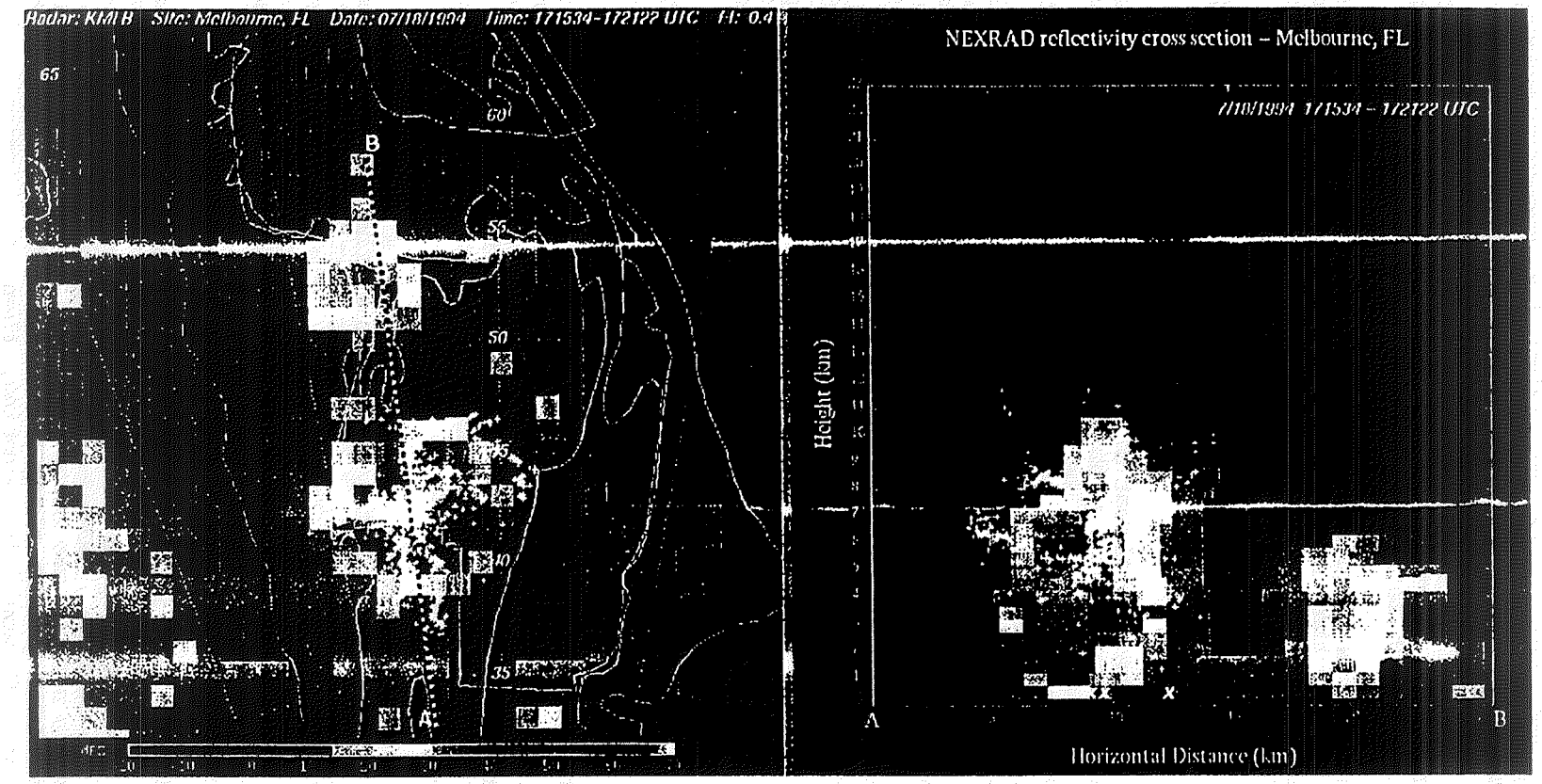


Figure 3-6. Radar and LDAR depiction of the storm from 1715-1721 UTC. Left, composite reflectivity with LDAR and LLP (black X) superimposed. Right, cross-section.

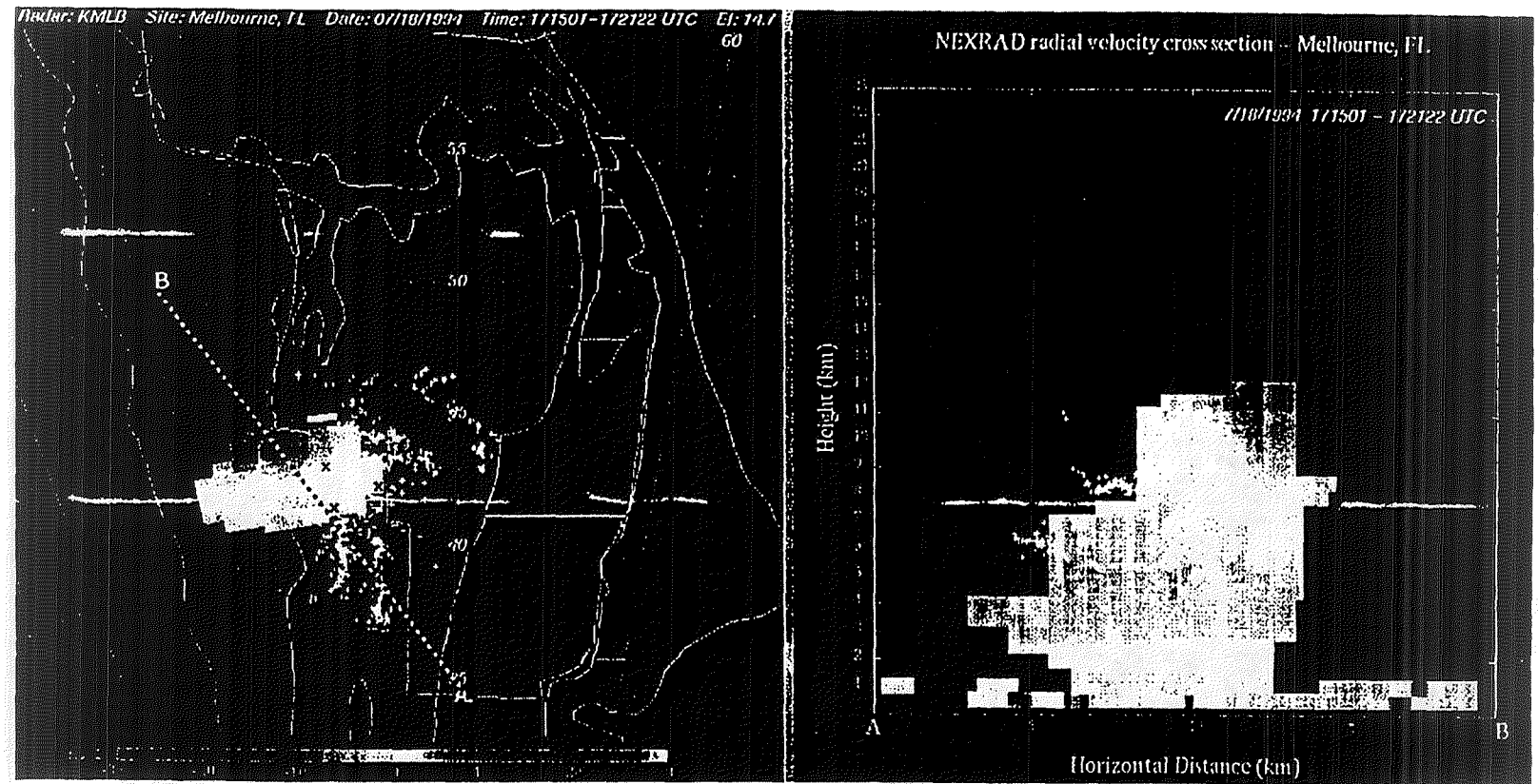


Figure 3-7. Doppler velocities in the storm of Fig. 3-6. Left, velocities in the scan at 14.7° elevation angle are positive (dark; northbound) in the northern part of the echo and negative (light; southbound) in the southern part, indicating divergence in the upper portion of the storm. Right, cross-section view.

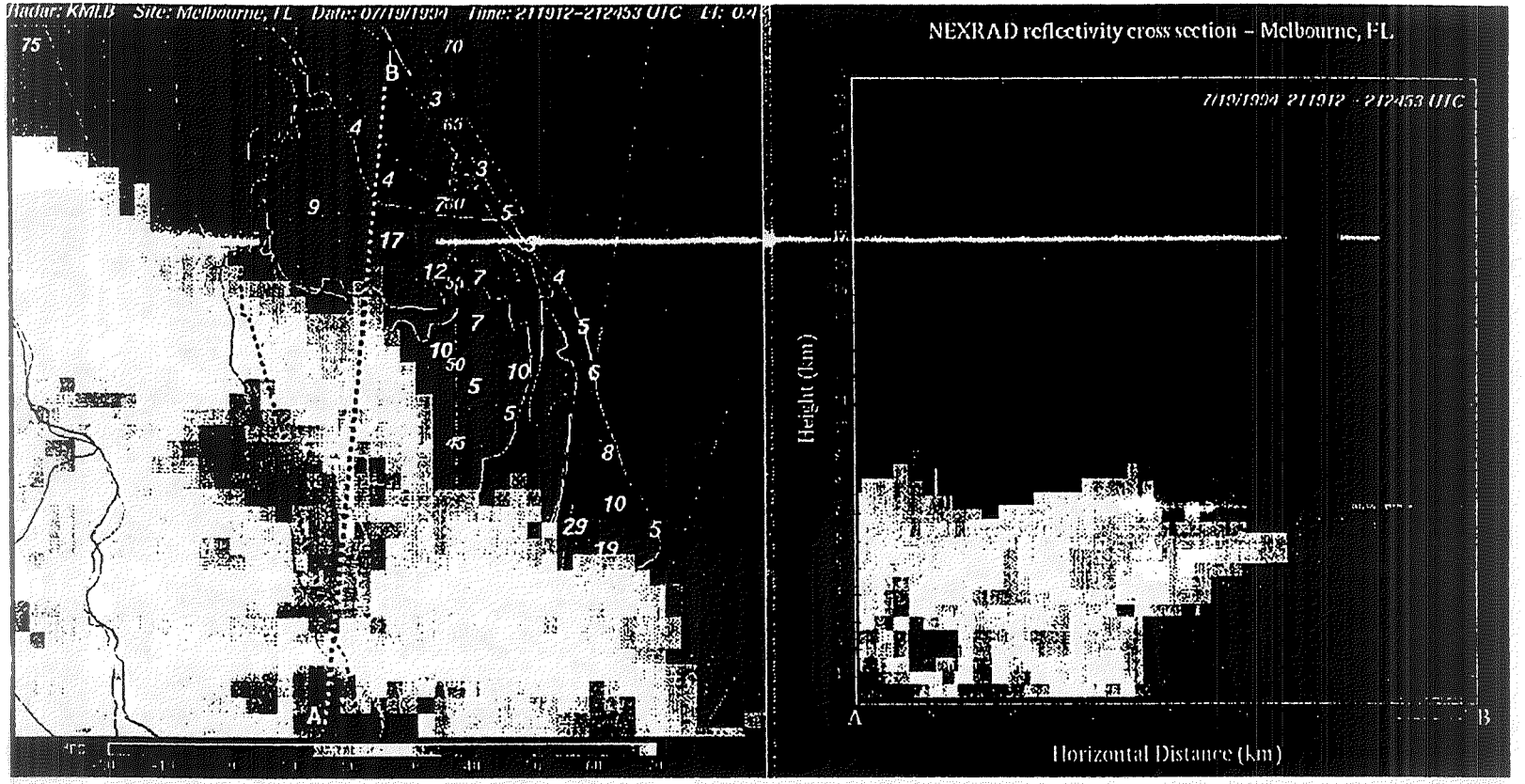


Figure 3-8. Radar depiction of the post-threat storm on 19 July 1994, from 2119-2125 UTC. Left, composite reflectivity and one-minute-average field mill (LPLWS) values from 2120, in tenths of kV/m. Right, cross-section.

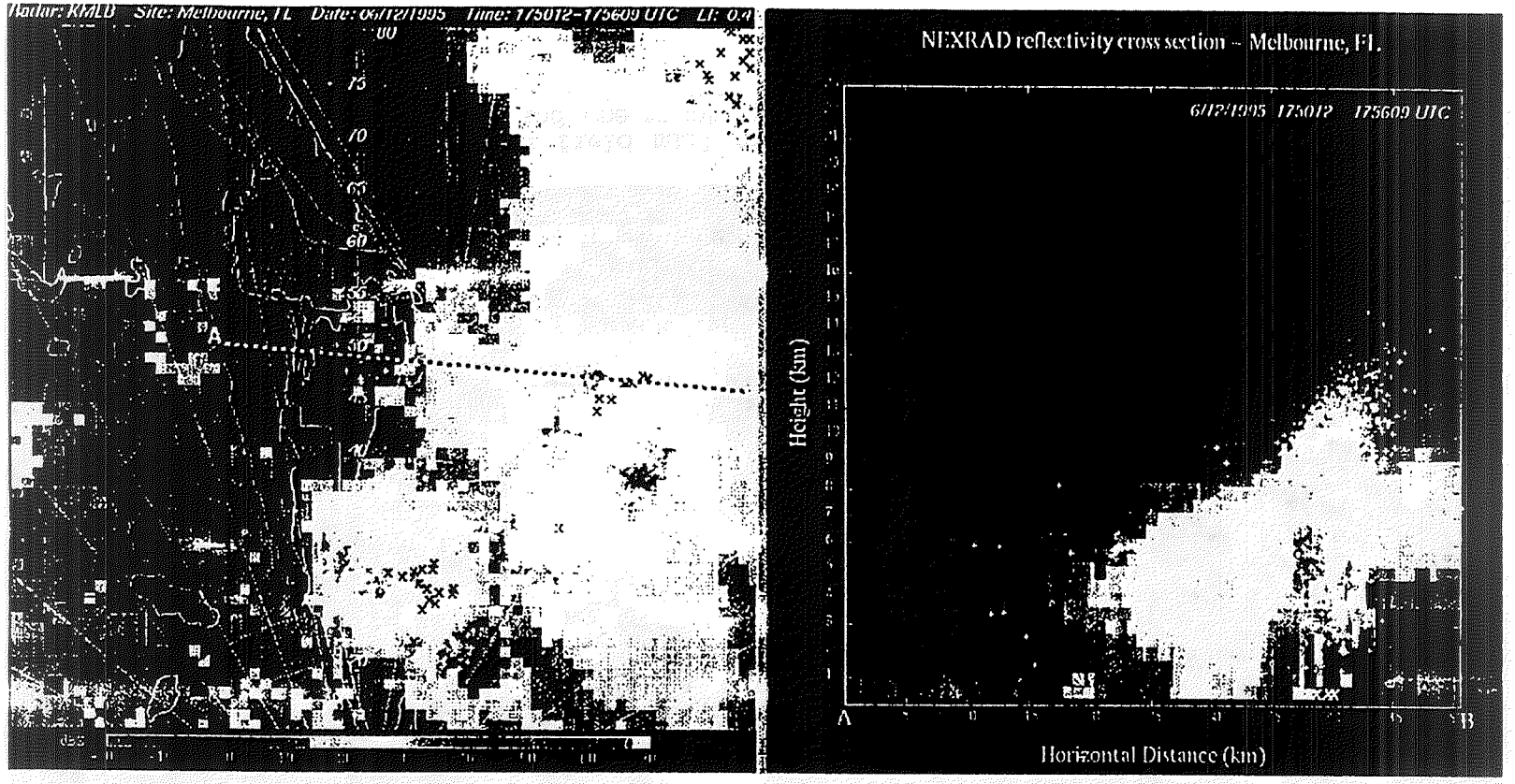


Figure 3-9. Radar and LDAR depiction of the post-threat storm on 12 June 1995, from 1750-1756 UTC. Left, composite reflectivity, LDAR (while + symbols), and LLP (black x). Right, cross-section.

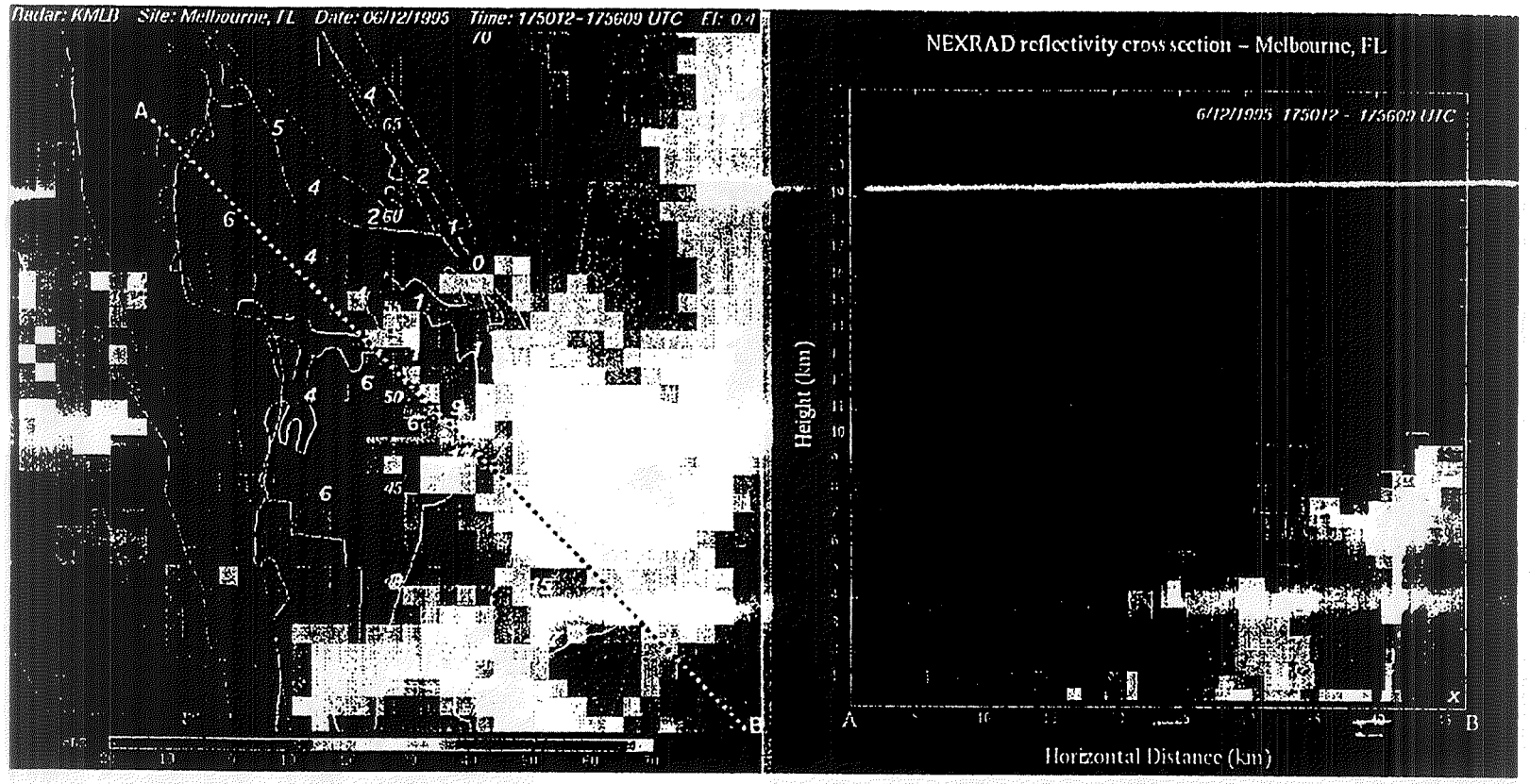


Figure 3-10. Radar and field mill data corresponding to Fig. 3-9. Left, composite reflectivity and one-minute-average field mill (LPLWS) values from 1755, in tenths of kV/m. Right, cross-section.

Figure 3-11 is from a day where lightning was from convective accentuations of more widespread layered or banded clouds. Use of maximum reflectivity as a crude lightning location predictor on this occasion would have given a very poor portrayal of the locations of peak lightning threat, as shown. Lightning on this occasion emanated from regions where sufficiently large reflectivity occurred at higher altitudes. On this day of more mesoscale rather than purely buoyant convective forcing, it appears that lightning was initiated where and when 30 dBZ echo occurred above the 7 km level. As shown in Fig. 3-11, the LDAR discharges then propagated along the top of the cloud and along the gradient atop the region of highest reflectivities. This resulted in primarily a downward slope, except near about 25 km on the section where a localized turret elevated cloud top for a short distance. Ground strikes were quite uncommon on this day, as only 5 ground strikes were indicated within 5 n.mi. of any warning site using both LDAR and LLP to detect ground strikes.

Figure 3-12 shows another LDAR discharge pattern on this occasion, and the only ground strike indicated by both LDAR and LLP. Once again the LDAR discharge pattern appears to have initiated not from the region of highest reflectivity but from the region of highest reflectivity above a critical altitude, at position near 10 km along the cross section. Use of the hypothesized 20 dBZ edge rule would have been difficult in this case, as there is so much low-level echo above 20 dBZ. Further, this case points out a potential danger in the inference drawn regarding Fig. 3-10. In that case a lower layer of echo > 20 dBZ was ignored because it was not the one in which LDAR discharges were occurring. In the case of Fig. 3-12, an echo from low-levels protruded upward into the base of the LDAR layer, perhaps not coincidentally becoming the site for a ground strike. Clearly additional study is needed to clarify the circumstances under which ground strikes occur beneath primarily layered LDAR discharges.

Figure 3-13 shows echo from the same case at a later stage. All ground strikes affecting any warning site have ended as of 2323, though field mill readings remain very large. A working hypothesis in this case might have been that the ground strike threat would cease once 30 dBZ echo no longer reached the 7 km level. Additional study of the meteorological bases for phase 2 termination is needed.

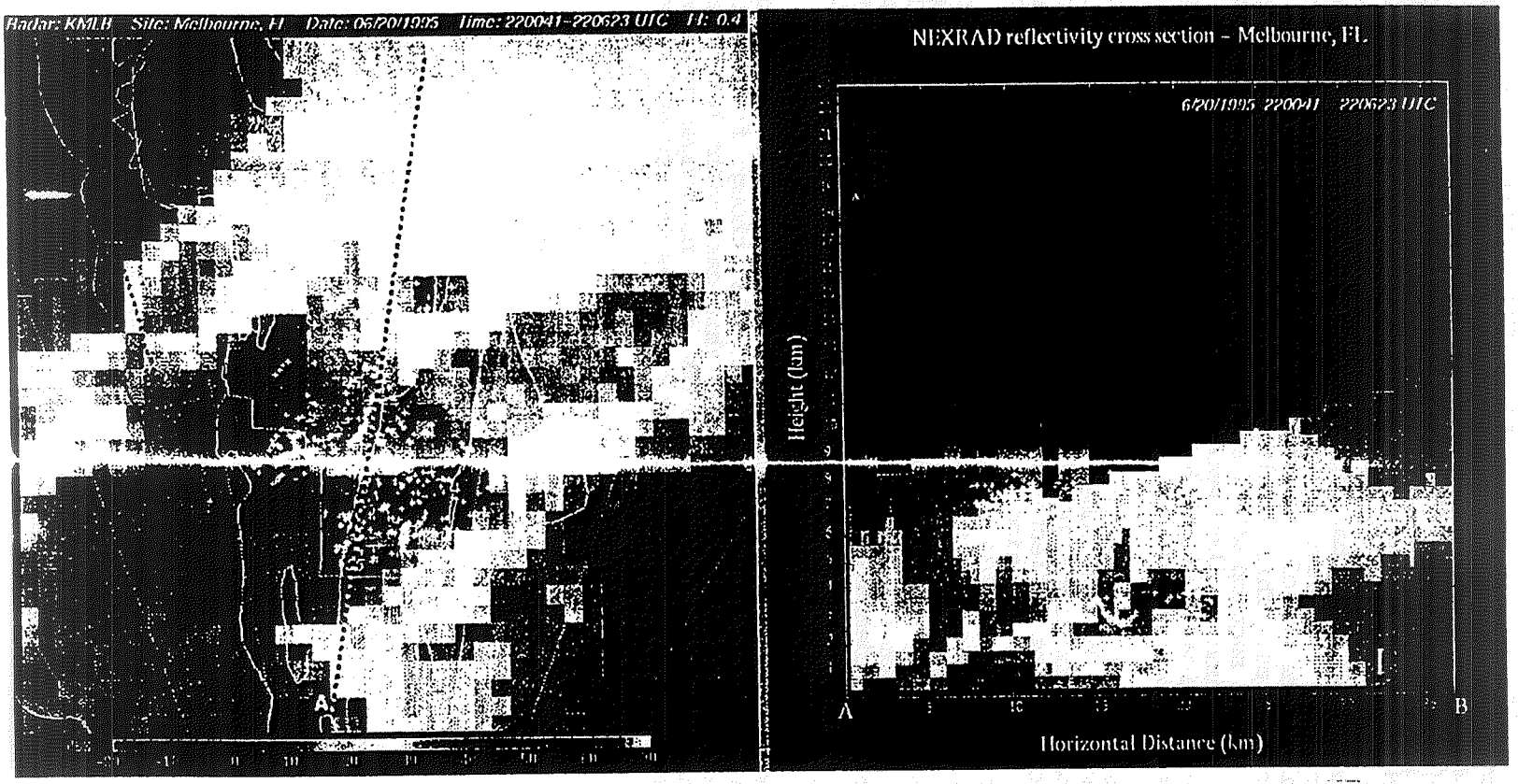


Figure 3-11. Radar and LDAR depiction of quasi-horizontal lightning layers on 20 June 1995 from 2201-2206. Left, composite reflectivity and LDAR. Right, cross-section.

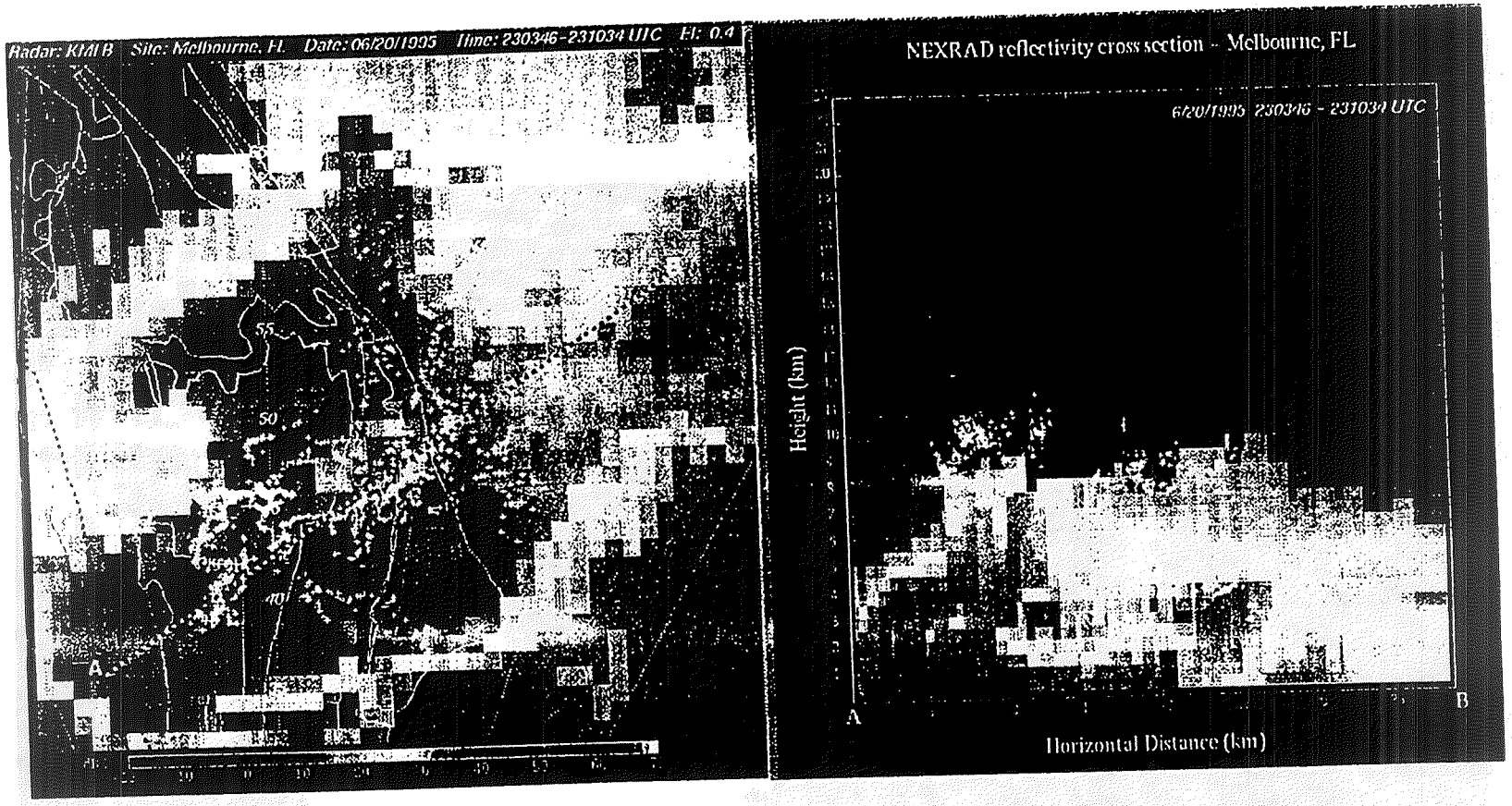


Figure 3-12. Radar and LDAR depiction of this case, including a ground strike, from 20 June 1995, 2304-2310 UTC.

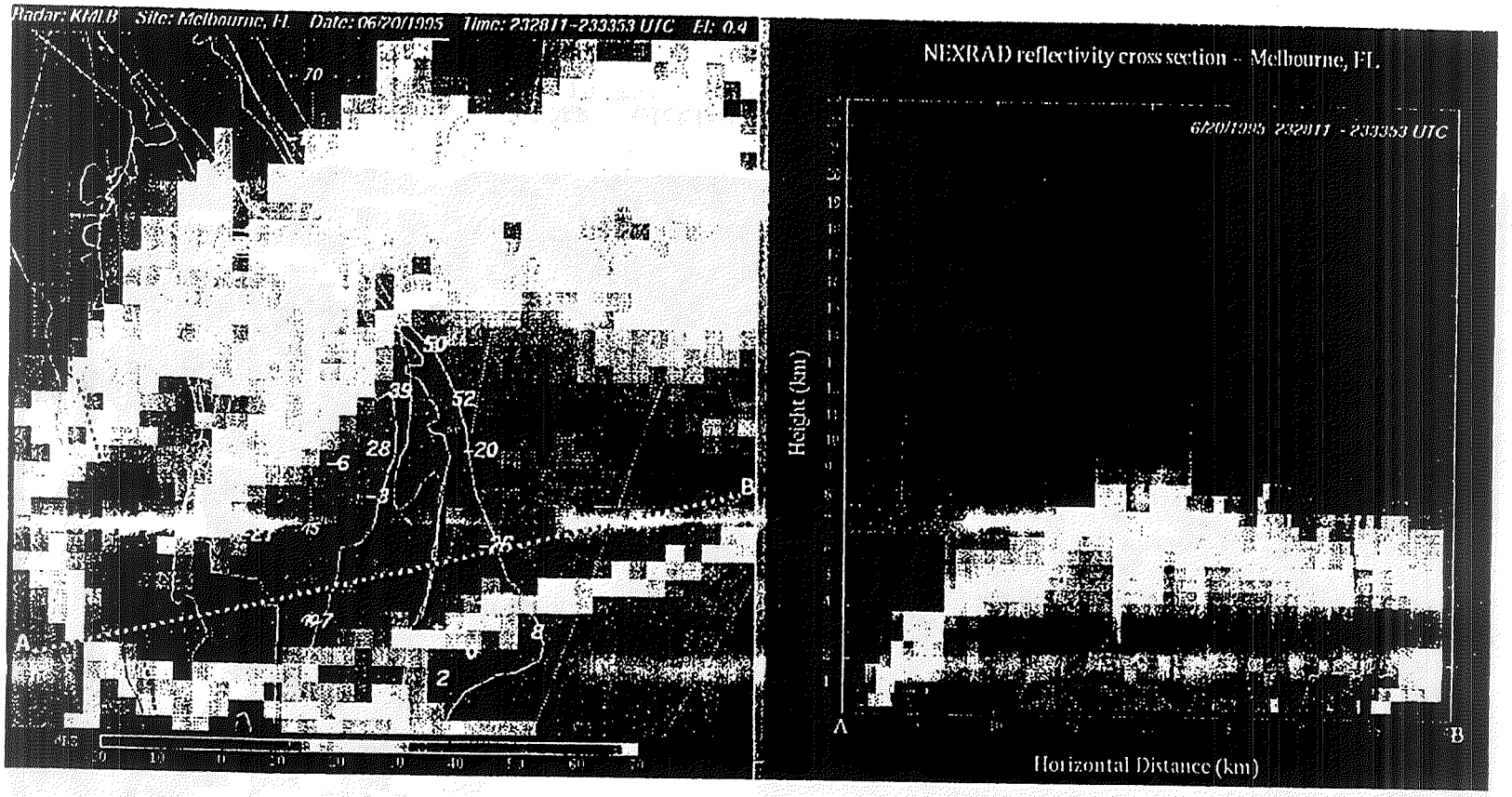


Figure 3-13 Radar and field mill depiction of the storm at a later stage, at the time of the last ground strike that affected any warning area. From 20 June 1995, 2328-2334 UTC. Left, composite reflectivity and one-minute-average field mill (LPLWS) readings from 2331, in tenths of kV/m. Right, cross-section.

V. CONCLUDING REMARKS

A strong case can be made for utilization of all available meteorological data sets in the forecasting of lightning. In particular, the forecasting of the in situ development of thunderstorms over the warning sites requires information beyond LDAR and LLP data, which only indicate when lightning has developed. Field mills sometimes provide advance information in these cases, but the use of radar, satellite, and surface wind tower mesonet data is also crucial. Likewise, it is difficult to be sure that lightning ground strikes have ended without the use of field mill and especially radar data to augment the LDAR and LLP data.

To most effectively use the various data sets, they should be overlaid on radar or satellite imagery. We have developed a software package that allows LDAR, LLP, field mill, and surface data to be overlaid onto radar imagery, as illustrated in this report. The software package produces plan and cross-section views of the radar and overlaid data sets. It can also produce maps of composite reflectivity (largest value in the column above each point), storm tops, and vertically integrated liquid water content (VIL). Additionally, we have developed software to implement the Velocity-Azimuth-Display (VAD) method for generating time-height series of mean wind, deformation, divergence and vertical velocity above the radar. The software package currently runs with archived data sets, but is expected to be adaptable for real-time use.

We have used the display software to examine 10 cases from June and July of 1994 and 1995. Some preliminary conclusions have been drawn, but they should be examined in more detail with further case studies, or examined on a real-time basis.

1. Radar, satellite, and surface mesonet data can help to pinpoint the location of thunderstorm formation. Mesoscale boundary detection, through use of satellite and radar imagery, streamline and convergence plots, temperature and temperature change plots, and vector wind change plots can yield valuable clues.
2. Lightning did not first appear in most developing thunderstorms until a sufficient mass of precipitation had developed at levels where temperatures at least as cold as -15 to -20°C would allow mixed-phase electrification processes to operate. Thresholds suggested to date are 30 dBZ or greater at altitudes of at least 8 km. Storm-top divergence should be monitored to help anticipate whether storms approaching this threshold are still in a growth stage (still divergent aloft).
3. In one case of small-scale convection embedded within larger-scale forcing, attaining 30 dBZ at 7 km appeared to be sufficient to initiate lightning.
4. The end-of-threat lightning forecast appears to require just as much, if not more, meteorological judgement. In most cases lightning did not strike ground more than about 2 km outside regions of the storm where low-level (0-4 km) reflectivities fell below 20 dBZ. However, there were notable exceptions.
5. Field mill readings typically remained above 1 kV/m tens of minutes or more after ground strikes had ended at a site, due to layered cloud left behind after the thunderstorm core had passed. Additional study is needed to determine joint meteorological conditions that enable forecasts of lightning threat cessation that are maximally safe and yet not overly cautious.

REFERENCES

1. Maier, M.W., A.G. Boulanger, and R.I. Sax: An initial assessment of flash density and peak current characteristics of lightning flashes to ground in South Florida. NUREG/CR-1024, U.S. Nuclear Regulatory Commission, 1979.

2. MacGorman, D.R., M.W. Maier, W.D. Rust: Lightning strike density for the contiguous United States from thunderstorm duration records. NUREG/CR-3759, U.S. Nuclear Regulatory Commission, 1984.

3. Maier, L.M., E.P. Krider, and M.W. Maier: Average diurnal variation of summer lightning over the Florida Peninsula. *Mon. Weather Rev., 112*(6), June, 1984, pp. 1134-40.

4. Krider, E.P., A.E. Pifer and D.L. Vance: Lightning direction-finding system for forest fire detection. *Bull. Amer. Meteorological Soc., 61*, 1980, pp. 980-86.

5. Lennon, C. and L. Maier: Lightning mapping system. NASA CP-3106, Vol, II, 1991 International Aerospace and Ground Conference on Lightning and Static Electricity, 1991, pp. 89-1 to 89-10.

6. Forbes, G.S.: Lightning studies using LDAR and LLP data and applications to weather forecasting at KSC. 1993 Research Reports, NASA/ASEE Summer Faculty Fellowship Program (C.R. Hosler, C. Valdes, T. Brown, eds.), NASA CR-194678, Grant NGT-60002 Supplement 11, 1993, pp. 165-194.

7. Forbes, G.S.: Lightning studies using LDAR and companion data sets. 1994 Research Reports, NASA/ASEE Summer Faculty Fellowship Program (L.A. Anderson, E.R. Hosler, W.Camp, eds.), NASA CR-197448, pp. 205-233.

8. Holle, R.L., A.I. Watson, R.E. Lopez, R. Ortiz: Meteorological aspects of cloud-to-ground lightning in the Kennedy Space Center region. Preprints, 26th AIAA Aerospace Sci. Meeting, Reno, NV, 1988, 12pp.

1995 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

S8-32
7748
P. 26
(26)

EVALUATION OF WIRELESS DATA COMMUNICATIONS AT KSC

Dr. Raghvendra R. Gejji
Associate Professor
Electrical and Computer Engineering Department
Western Michigan University
Kalamazoo, Michigan

KSC Colleague - Ray Pecaut
Communications/Networks

Contract Number NASA-NGT-60002
Supplement 19

August 9, 1995

ACKNOWLEDGEMENTS

I thank Dr. Ray Hosler and Ms. Kari Stiles of U.C.F., and Jerry Barnes of NASA for their support in the Summer Faculty Fellowship Program. I owe a huge debt of gratitude to my NASA colleague Ray Pecaut and Rick Birr of INET for their constant cooperation and encouragement. Thanks are also due to Mike Bertucci, John Farrell, Jim Hillis, Kirk Bigelow and Dave Wedekind for their support.

ABSTRACT

This project is motivated by the need for temporary or emergency provisioning of LAN service at KSC. The main goal of the project was to evaluate existing wireless bridge equipment in the KSC environment. Wireless bridge equipment can be used to make a wireless connection between two remotely located LAN segments. This report describes the experimental setup used to evaluate the equipment, including antenna connections, workstation connections, bridge software and workstation software. The rangefinder program on the bridge was used to gather data about how the RF propagation environment at KSC affects the performance of the wireless bridge. Data was gathered for indoor as well as outdoor propagation. The report concludes with recommendations on how to take into account the particular terrain and building structures at KSC to design future applications of wireless bridges.

SUMMARY

This project is motivated by the need for temporary or emergency provisioning of LAN service at KSC. There are two ways that wireless technology may be used in a LAN environment. One is to provide a wireless connection from a single terminal to a remote wired network. The other is a bridge connection between two remotely located network segments. In this project we will only be discussing wireless bridge equipment. This can address situations such as emergency disruption of LAN connectivity to a certain key location, or the need for LAN connectivity to temporary structures such as trailers. Additionally, wireless LAN technology can provide flexibility in managing network topology and reduce labor costs for cable installation in situations where the arrangement of work areas is subject to ongoing change in response to changing projects, personnel and equipment. The wireless bridge equipment available at KSC is the INTERSECT(TM) bridge made by PERSOFT Inc., of Madison, Wisconsin. The main goal of this project was to evaluate the capabilities of this existing equipment in the KSC environment.

The experimental setup consists of two remote sites linked by wireless transmission. At each site, the setup consists of a wireless bridge, antenna, and workstation. The antenna can be indoor or outdoor depending on whether the application calls for in building or out of building transmission. The indoor antenna is omni-directional patch antenna about six inches in size. The outdoor antenna is 4-element Yagi antenna, approximately 2 feet, or a 10-element Yagi, approximately 4 feet. If an outdoor antenna is used, it needs to be mounted on a proper mast, taking into account weather and safety considerations. Normally many workstations may be connected to the bridge at each site. In our experiment, only one workstation was connected to each bridge using UTP null-modem cable with RJ-45 connectors.

The bridge comes equipped with a range finder program that can be used to assess the quality of the radio link. Other features of the bridge software are the bridge configuration program which can be used to set up traffic screening and spanning tree parameters for traffic routing. These latter features were not used in our project. The workstations used were UNIX workstations with ftp (file transfer protocol) software. This allowed remote login, remote directory, and two-way file transfer between the workstations over the wireless link for demonstration purposes. Wireless file transfer was demonstrated indoors from the East end of the EDL building to the West end, and outdoors between EDL and CIF buildings.

The rangefinder software on the bridges was used to gather RF propagation data under various application situations, both indoors and outdoors. Data was gathered for indoor propagation and link quality at many locations throughout EDL building. Data for outdoor propagation was gathered by setting up one site in the EDL building, and the other site in a mobile van. Outdoor propagation data was gathered throughout the KSC industrial area, as far West as the KARS II park.

The propagation data show that equipment performance and radio link quality are seriously affected by physical structures that affect RF propagation. For outdoor propagation, these include high terrain, tall buildings, trees etc. For indoor propagation, the factors are, number of walls and construction materials, and propagation between floors. Plaster walls are quite forgiving, but masonry walls cause significant signal loss .

The final conclusion of the project is that the PERSOFT equipment is suitable for use at KSC, but the performance for any given application depends on the physical RF propagation

environment, which needs to be carefully considered. Guidelines for carrying out an evaluation of the propagation environment are included in this report.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

CIF	Central Instrumentation Facility
EDL	Engineering Development Laboratory
FCC	Federal Communications Commission
FTP	File Transfer Protocol
ISM	Instrumentation, Scientific and Medical
KSC	Kennedy Space Center
LAN	Local Area Network
TCP/IP	Transmission Control Protocol/Internet Protocol
UTP	Unshielded Twisted Pair

I. INTRODUCTION

This project is motivated by the need for temporary or emergency provisioning of LAN service at KSC. This can address situations such as emergency disruption of LAN connectivity to a certain key location, or the need for LAN connectivity to temporary structures such as trailers. Additionally, wireless LAN technology can provide flexibility in managing network topology and reduce labor costs for cable installation in situations where the arrangement of work areas is subject to ongoing change in response to changing projects, personnel and equipment [1].

There are two ways that wireless technology may be used in a LAN environment. One is to provide a wireless connection from a single terminal to a remote wired network. The other is a bridge connection between two remotely located network segments. In this project we will only be discussing wireless bridge equipment.

The wireless bridge equipment available at KSC is the INTERSECT(TM) bridge made by PERSOFT Inc., of Madison, Wisconsin. The main goal of this project was to evaluate the capabilities of this existing equipment in the KSC environment. The nominal capabilities of the Intersect bridge are delivery of data at the rate of 2 Mb/s between remote locations upto 3 miles apart outdoor, or upto 800 feet indoors. The equipment uses 26 MHz of bandwidth in the FCC unlicensed ISM band (902-928 MHz). Spread spectrum technology is used to keep the transmitted power to only 250 mW. The receiver sensitivity is -72 dBm.

II. EXPERIMENTAL SETUP

2.1 Bridge connections

The Intersect remote bridge connections are first briefly described. Please refer to the User manual [2] for a more detailed description. There is one connector for the antenna cable. This connector is F-type. The antenna cable is 75-Ohm RG-6/U type cable. Each bridge has the function of forwarding traffic between a wired Ethernet segment and the wireless remote connection. There are three possible ways of connecting the Intersect bridge to the wired Ethernet segment. The default method is Unshielded Twisted Pair (UTP) cable to the hub of a standard 10BaseT ethernet. An RJ-45 port is provided for this. Other Ethernet adapter ports are provided for alternate connections to the Ethernet. An AUI port is available for connecting to 10Base5 standard equipment, also called thick ethernet. If this method is used an AUI drop cable and a transceiver would be required. A third alternative is available to interface to thinwire ethernet, also called 10Base2 or coaxial cable equipment. For this a BNC port is provided on the bridge. A coaxial cable would have to be connected from the BNC port to a T-connector on the ethernet cable. The network connection used in our experiment was a modified UTP connection, as will be described later.

2.2 Antenna Connections

Each bridge is supplied with two types of antennas. One is a flat antenna about 6 inch size that can be mounted on the wall and is for use in indoor wireless applications. The other is a 4-element Yagi antenna about 18 inch size, and is for outdoor applications. The wall-mount antenna is supplied with its own 5-feet cable with F-type termination to connect directly to the bridge. The outdoor antenna is supplied with an N-type connector. A separate 33 feet cable is required. This cable should be 75-Ohm cable with N-type termination on one end to connect to the antenna, and an F-type male termination on the other end to connect to the bridge.

2.3 Workstation connections

Any type of workstation with ethernet interface can be used to connect to the bridge. For normal operation, both the workstation as well as the bridge would be connected to a hub of the ethernet LAN segment. In our case, for demonstration purposes, it was decided that only one workstation would be used on each side of the wireless connection, instead of a LAN segment. Since there is only one workstation to be connected to the bridge, the hub can be eliminated, and the workstation can be connected directly to the bridge by UTP cable. But if this is done, it is necessary to cross-connect the receive and transmit terminals at the two ends of the cable. Such a cable is called a null-modem cable. A diagram showing the RJ-45 terminal pin connections that should be cross-connected is shown in Fig. 2.1.

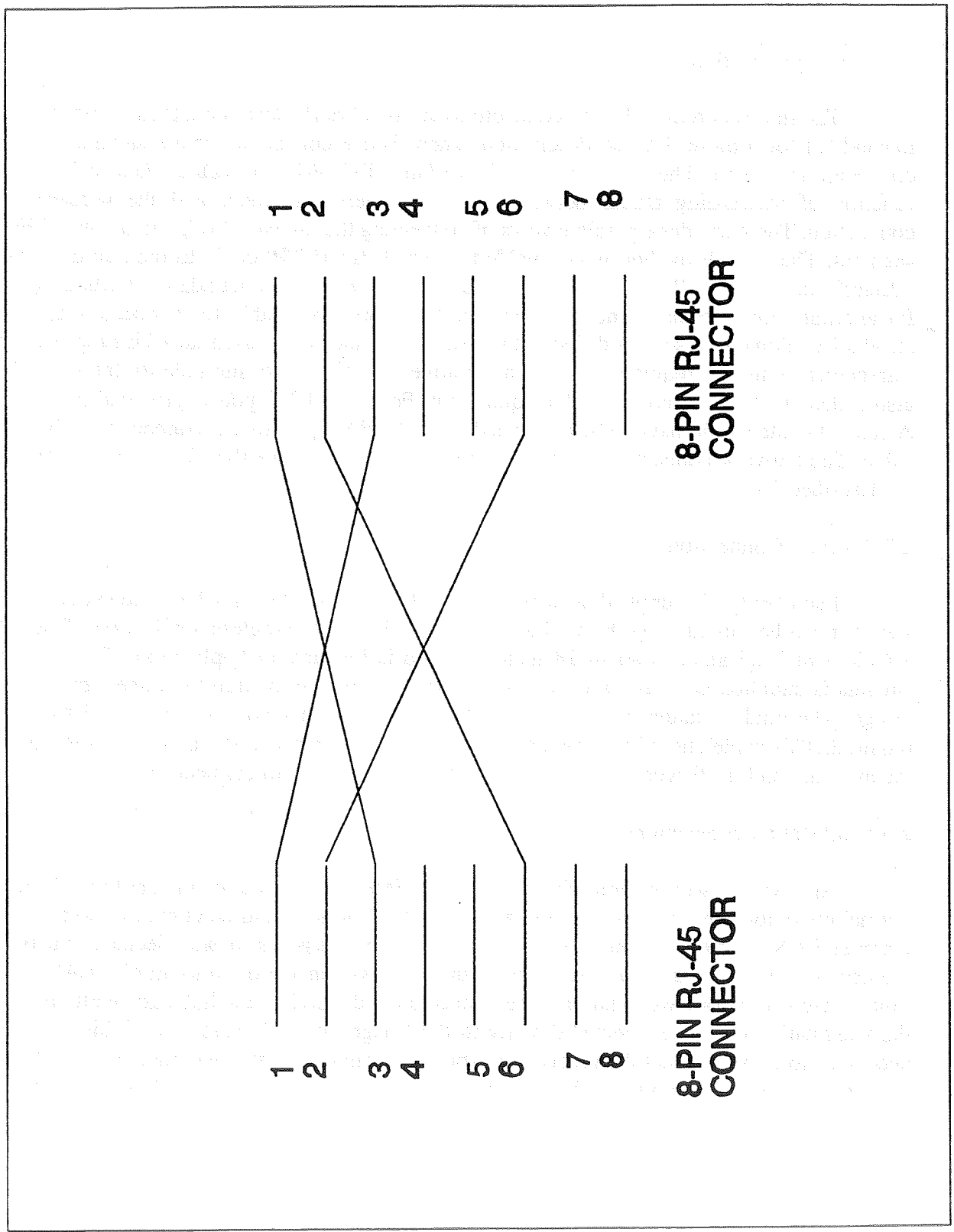


Figure 2.1 RJ-45 Connections for null-modem cable

2.4 Experimental setup

Fig. 2.2 shows the overall schematic of the experimental setup, including bridges, antennas and workstations. The connections between the elements of the setup were made as described above. Note that antenna may be an indoor antenna or an outdoor antenna.

2.5 Outdoor Antenna Installation

Outdoor antennas should be installed using manufacturer's recommended safety precautions.The antennas should be grounded to protect against lightning. Accidental contact with power lines during installation can cause injury or death. More complete instructions on safety precautions for roof-antenna installation are available from the Consumer Safety Commission. For this project, outdoor antennas were installed on CIF and EDL buildings, a distance of approximately 1 mile. A 10 element Yagi antenna supplied by Persoft was installed on the roof of the EDL building using U-bolt mounting to a vertical mast of 2 inch diameter aluminum pipe. It is important that mast height be chosen so as to allow line-of-sight access between the two antennas. Any intervening structures can cause problems with reception. A 75 feet long 75-Ohm cable was connected from the antenna to a window terminal in the building wall. The 30 foot long antenna cable supplied by Persoft was installed from the window terminal to the bridge terminal. In the CIF building, the 10-element Yagi antenna supplied by Persoft was installed on an existing antenna mast. An existing cable was used to connect to the bridge placed on the second floor of the CIF building. This cable is 200 feet long and is 50-Ohm rating. Antennas should be installed on both sides so that the antenna elements are vertical with the driven element (gray) up.

2.6 Software

The bridges come supplied with software on floppy diskette. The floppy diskette is a 3.5 inch, 720K diskette. For protection against accidental deletion, a copy of the diskette supplied should be made using any IBM-compatible PC workstation. Each bridge has its own unique floppy diskette. The two diskettes should not be interchanged.

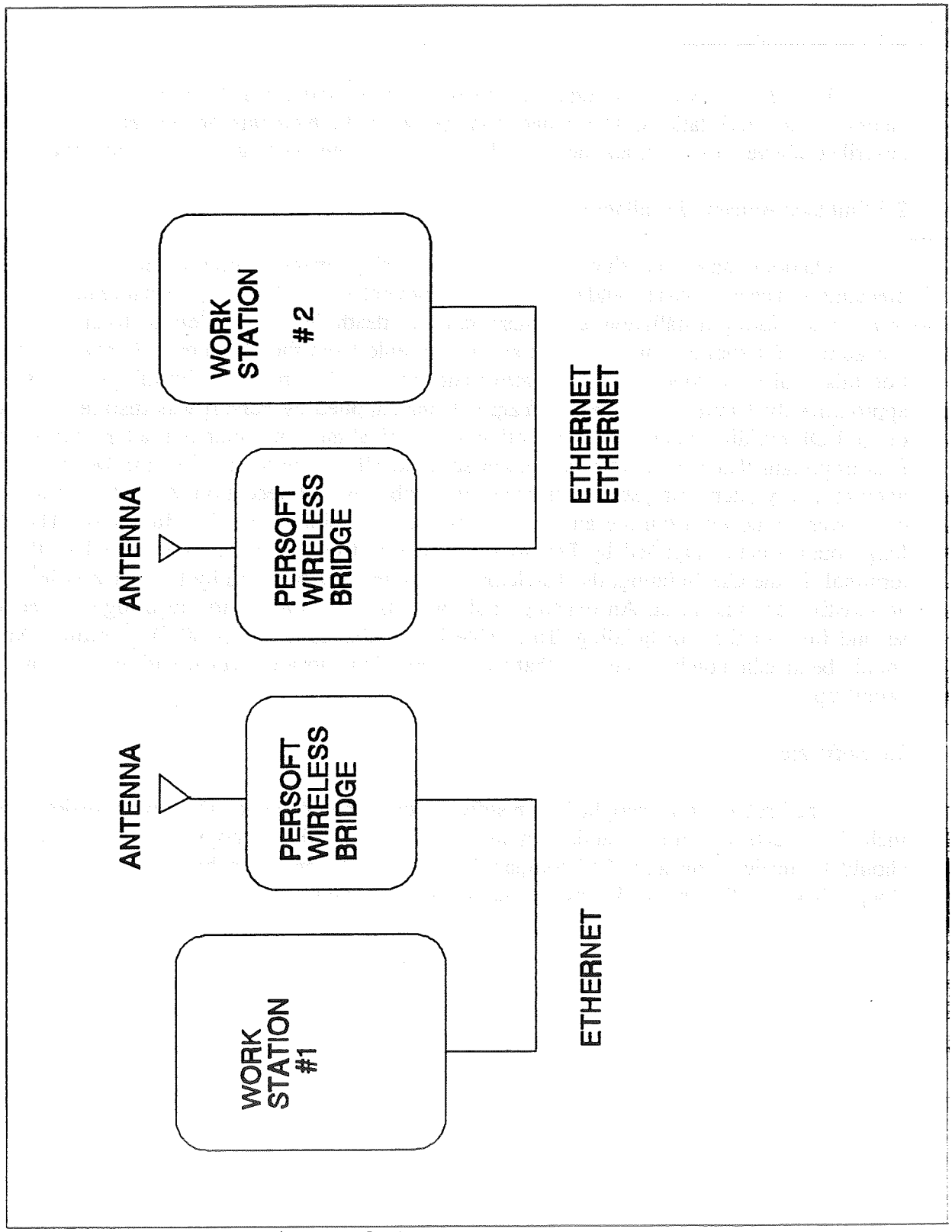


Figure 2.2 : Experimental Setup

2.7 Rangefinder software

When the bridge is powered up with the appropriate floppy diskette inserted in the floppy disk drive, a menu appears. Option no. 2 from the menu invokes a rangefinder program that can be used in measuring the quality of the radio link and in fine-tuning the antenna pointing. The rangefinder program transmits data packets between the two sites. Statistics of the link quality are displayed on the consoles at both ends. Thus it is possible to monitor the reception quality at both ends of the link from one end alone. The rangefinder program transmits a fixed 72 packets per second from each station. It displays the percentage of packets received without error. It also displays the signal quality, signal level and signal-to-noise ratio at both ends. The signal level can be from 0 to 36 dB. 18 dB is required for satisfactory operation. Signal-to-noise ratio can be from 40 to 80 dB. 60 dB is required for satisfactory operation. The signal quality factor can be 0 to 12. A quality factor of 7 is required for satisfactory operation.

2.8 Bridge configuration

Option no. 3 on the main menu of the bridge software allows customization of each bridge within its environment. The various parameters that can be set are briefly discussed. Each bridge is assigned a network id ranging from 0100 to 7FFF hexadecimal. All bridges of a given id will communicate with one another. This allows multiple pairs of remote bridges to operate on the same frequency band without interfering with one another. Other parameters that can be used to customize each bridge are spanning tree parameters and traffic screening parameters. Spanning tree parameters determine whether the Intersect bridge will become the root bridge or the preferred bridge in a situation where a loop of bridges develops [3]. The spanning tree parameters that can be chosen for each bridge are: Bridge id and priority, wireline port cost and wireless port cost. These parameters are more fully described in the references. Traffic screening parameters can be set to delimit the type of traffic that will be either forwarded or blocked by the bridge. Traffic to be forwarded or blocked can be identified by ethernet station id, or by TCP/IP traffic type.

2.9 Workstation network software

Many types of network software could be used for communicating between workstations such as ftp, email, etc. The network software takes application data and converts it to ethernet packets. In our experiment, the two workstations were both running Xenix operating system and were both supplied with ftp (file transfer protocol) program. With this program it is possible to login to the remote workstation, do a directory listing from the remote workstation, and transfer a file from remote to local (get command) or local to remote (put command). By measuring the time it takes to transfer a large file (approx. 1 Megabytes) we can get an idea of the link speed.

III. PROPAGATION TEST RESULTS

3.1 Indoor Network Test

The experimental setup described in Fig. 2 was set up in Room 140 of the EDL building. Remote login, remote directory and a file transfer from one workstation to the remote workstation was demonstrated to NASA personnel on June 7, 1995. During this demonstration two indoor antennas were used in close proximity. The rangefinder software showed that the signal level and quality were 12 and 80 dB respectively, the highest achievable values.

3.2 Indoor Propagation Test with Patch Antennas

For measuring radio propagation data, the two workstations were disconnected from the experimental setup of Fig. 2. Since the bridge generates its own traffic of Ethernet packets during the rangefinder test, a workstation is not necessary. Only a bridge and antenna are needed at each end of the link. One bridge and indoor antenna was left remaining in Room 140. The other bridge, along with monitor for display were placed on a cart. The rangefinder program was initialized at the stationary location. The site name was chosen as "Room 140." The screen then displayed the message "Waiting to synchronize with partner." At this point the mobile bridge was moved to a suitable location where AC power is available, and the rangefinder program was initiated. The site name was chosen as "mobile". At this point the two stations initiated wireless contact, and the monitor of the mobile station displayed signal statistics which were recorded. The rangefinder program was terminated at the mobile station. However, when the message "Notify partner to exit?" was displayed, the "no" option was chosen. Hence the stationary bridge in Room 140 remained in the original rangefinder mode, "waiting for partner to synchronize", even though the mobile station had exited the program and even powered down. The process was repeated at different locations around the EDL building, each time ensuring that the stationary bridge remained in the rangefinder mode. At each of the selected locations, signal level, quality, signal-to-noise ratio, and percentage of packets received correctly were recorded for both ends of the link. Table 3.1 shows the data obtained. The goal at this stage of the project was to test for successful link from one end of the EDL building to the other, i.e. from Room 140 on the West end of the first floor to Room 125 on the East end.

TABLE 3.1 : Indoor Propagation Data with Patch Antennas

Loc.	Dist.	Mobile				Room 140			
		Signal Qual.	S/N Ratio	Signal Level	% correct	Signal Qual.	S/N Ratio	Signal Level	% corr.
	feet		dB	dB			dB	dB	
Min. Req'd for proper operation		7	60	18		7	60	18	
Adjacent	0	12	80	35	100	12	80	35	100
Around corner	10	11	75	32	100	11	76	34	100
Outside Rm 128	200	10	50	10	98	10	50	10	98
Room 125	300	8	50	8	94	8	50	8	94

As the table shows, perfect reception was not achieved between Rooms 140 and 125, even though the distance is only 300 feet, and the equipment has a nominal range of 800 feet indoors. We will see later that this was due to two concrete walls that block the direct path of the radio wave between the two rooms. The walls are the West wall of the hallway adjacent to Room 110, and the East wall of Room 124. As a result perfect 100% correct reception was not possible, It should be noted that even when the percentage of correctly received packets is less than 100%, communication is still possible but at a rate proportionately reduced from the rated 2 Mb/s. It should also be noted that in this case since the straight line path was blocked by concrete walls, communication depends on radio waves reflected by building structures and traveling along hallways and through doors etc. This type of propagation mode is subject to so-called multipath effects, whereby drastic variation in signal strength is observed by small changes in location, antenna orientation etc. The table gives data for the best possible antenna location and orientation for a fixed workstation location. It was observed that in the presence of multipath, sometimes the best signal is obtained by pointing the antenna toward a corner facing the door, as if the radio wave enters the room through the door, is reflected from a corner and then picked up by the antenna. The best reception with the most range is possible when there is an unobstructed straight line path for radio waves between the two antennas. It was seen that the signal propagated through plasterboard without much loss, but was unable to propagate through masonry concrete walls. Notice that the loss characteristic of the radio signal through various materials is strongly dependent on the frequency of transmission used.

3.3 Indoor Propagation With One Directional Antenna

Since the first indoor test was not satisfactory, another test was conducted. This time the small omnidirectional patch antenna on the mobile site was replaced by a 4-element directional Yagi antenna. It was hoped that the increased antenna gain would improve performance so 100% of the packets would be correctly received. Please note that in taking the following data, the bridge in Room 140 was connected to an omnidirectional patch antenna, while the bridge on the mobile station was connected to a 4-element directional Yagi antenna. The data is presented in Table 3.2. For each location noted in the table, the data reflects the reading obtained with the best possible antenna location and orientation within a 5 foot radius of the fixed bridge location. It was observed that when a direct straight line unobstructed signal path was available, small changes in antenna location and orientation were not significant. But where the direct path was blocked by masonry or concrete wall, the received signal exhibited severe multipath effects, and in these cases, antenna location and orientation were very important for obtaining the best signal. All the readings were taken in the main hallway outside the referred room, except for Room 125, where an inside reading was also taken.

Table 3.2 : Indoor Propagation Data with One Yagi Antenna

Loc.	Dist.	Mobile				Room 140			
		Signal Qual.	S/N Ratio	Signal Level	% correct	Signal Qual.	S/N Ratio	Signal Level	% corr.
	feet		dB	dB			dB	dB	
Min. Req'd for proper operation		7	60	18		7	60	18	
Rm 144	10	11	75	32	100	11	75	32	100
Outside Rm102B	35	11	70	26	100	11	66	26	100
O.S. Rm 107	75	12	60	15	100	12	58	16	100
O.S. Rm 150	90	11	57	13	100	11	56	16	100
O.S. Rm 122[1]	140	11	52	8	100	10	53	11	94
O.S. Rm126[1]	200	10	51	7	100	9	51	9	100
O.S. Rm170[1]	230	9	50	6	100	9	50	8	97
Inside Rm170[2]	230	11	53	11	100	11	53	13	100
Rm198 N-E	150	9	54	11	100	8	55	14	100
Rm198 South	150	9	56	12	100	10	55	14	100
O.S. Rm125[1]	300	9	50	6	55-100	9	50	9	40-100
Rm 125 West[1]	300	6	49	6	0	6	50	8	0

Notes:

(1) The reception quality was very sensitive to antenna position and orientation, suggesting the existence of severe multipath. The best results were obtained with the antenna directed toward an arbitrary direction not necessarily toward the other site, suggesting that we were picking up a reflection from a building structural element.

(2) It is interesting that the signal quality is poor just outside the door to Room 170, and is acceptable just inside. This is because the location just inside Room 170 is outside the "shadow" created by the concrete walls that are on the West side of Room 110, and on the East side of Room 124.

After the above data were taken, the two workstations were connected to the bridges and an ftp connection was demonstrated between Room 140 to a location just outside the door to Room 125.

3.4 Outdoor Propagation Data

Outdoor propagation data are presented in Table 3.3. These data were taken on June 23, 1995, using a mobile studio van.The conditions were wet and cloudy. A 6 dB Yagi antenna with 50 feet of 50-Ohm cable was mounted on the van. The van has the capability to raise, lower and rotate the antenna. It can also supply AC power for the bridge. The stationary site was established on the 2nd floor at the West end of the EDL building. A 10-element Yagi antenna was mounted on the roof of the EDL building. It was connected outdoors to an existing window terminal by a 75 foot length of 75-Ohm antenna. The inside of the window terminal was connected to the bridge by the Persoft-supplied 30 foot 75-Ohm cable. The antenna was directed West toward the CIF building. Data were taken by driving the van in a Westward direction.

Table 3.3 : Outdoor Propagation Data

Loc.	Dist.	Mobile				EDL			
		Signal Qual.	S/N Ratio	Signal Level	% correct	Signal Qual.	S/N Ratio	Signal Level	% corr.
	feet		dB	dB			dB	dB	
Min. Req'd for proper operation		7	60	18		7	60	18	
EDL Lot (Ant. up)	100	12	75	24	100	12	75	24	100
Robotics Trailer	600	12	61	21	100	12	65	20	100
O&C Tank	1200	12	64	23	100	12	65	20	100
O&C lot West	1800	12	61	22	100	12	63	19	100
H.Q. Lot East	2200	12	60	20	100	12	63	18	100
H.Q. lot West	3200	12	59	17	100	12	60	16	100
C Ave	3700	12	57	15	100	12	57	13	100
CIF lot (Xfrmr) See note[1]	4400	12	54	10	100	12	52	8	100
CIF lot (Ant. up 30')	4400	12	55	13	100	12	56	12	100
A Ave.	6500	12	56	14	100	12	56	12	100
Guard gate	6000	No	signal	O&C	sha-				
NASA Blvd at S.R. 3	7500	No	signal	due to	trees	sha-			
						dow			

KARS II East	7600	12	51	8	100	12	50	7	98
KARS II West	8400	11	48	6	0	8	47	4	0
KARS II W. (ant. up)	8400	12	52	8	100	12	51	7	100
Cause-way bet. SR3 and Space port	9300	No	signal	due to	sha-dow				
Space port	1100 0	No	signal						

Notes:
1. By rotating the antenna in this location it was seen that good signal quality is maintained over a 90 degree range.

From the above data it is seen that although the equipment is rated for 3 mile range, in practice the range is greatly affected by the shadowing effects of buildings, trees, and by land features such as high and low spots. We note that in the KARS II location, no signal was obtained with the antenna lowered. But when the antenna was raised so it cleared the treetops, then a good signal was obtained.

3.5 EDL to CIF Networking Test

For this demonstration one 10-element Yagi antenna was installed on the roof of EDL and CIF buildings each. In the EDL building the antenna was connected to a window terminal (West end of second floor) with a 75 foot long 75-Ohm cable. Another 30 foot long 75-Ohm cable (RG-6/U) was connected from the window terminal to the bridge . This type of cable is rated for a loss of 6-10 dB per 100 feet depending on the manufacturer. In the CIF building an existing 200 foot long run of 50-Ohm cable was used to connect from the roof antenna to the bridge located in a second floor lab. When both bridges were powered up, they automatically enter normal bridge operation mode in 30 seconds. It turned out that it was difficult to find suitable antenna positions on the two buildings that provided unobstructed line-of-sight view from one antenna to the other. The water tower on the O&C building creates an obstruction. Because of this the mast on the EDL building used for the outdoor propagation data proved to be unsuitable for the networking demonstration. The 8 foot mast was replaced by a 10 foot mast, and the demonstration was conducted for NASA personnel on July 6, 1995.

When operating in its normal mode, the bridge console, if connected displays traffic statistics in both directions. This data was not recorded for this project. A workstation was connected to the bridge at each end, and an ftp connection was established between the two workstations. A file of approximately 1 MByte size was transferred, and the process took about 30 seconds. This indicates a transfer rate of about 250Kb/s, which is less than the rated data rate of 2 Mb/s. The reason for this is that with only one workstation transmitting at a time, the data rate is limited by the disk access time needed to access the file to be transferred. In order to load the bridge to a level of 2 Mb/s traffic, it will be necessary to have multiple workstations on each end sending data simultaneously. This experiment was not conducted during this project.

3.6 Indoor Propagation Data with Two Directional Antennas

For these data, directional antennas were used on both sites. The stationary site was again set up in Room 140 of the EDL building, with the antenna directed East, so that the signal would propagate down the main hallway. Most of the readings were taken along this direction. For the last few readings, beginning with location no. 17, the antenna in Room 140 was reaimed to point toward the mobile. The mobile was equipped with a battery backed UPS (Uninterruptible Power Supply) that allowed readings to be taken without turning power off and outdoors. The data is shown in Table 3.4.

Table 3.4 : Indoor Propagation Data with Two Yagis

Loc.	Dist.	Mobile				Room 140			
		Signal Qual.	S/N Ratio	Signal Level	% correct	Signal Qual.	S/N Ratio	Signal Level	% corr.
	feet		dB	dB			dB	dB	
Min. Req'd for proper operation		7	60	18		7	60	18	
1	275	12	51	9	100	12	52	11	100
2[1]	140	12	58	14	100	12	58	18	100
3	220	9	51	6	30	7	51	9	38
4	230	7	51	6	0	12	54	9	0
5	320	9	50	7	0	10	50	8	0
6	310	12	52	12	100	12	53	13	100
7	300	8	49	7	0	9	53	11	0
8	200	8	49	6	5	7	49	8	0
9	190	11	52	9	91	10	52	11	100
10	160	12	54	13	100	12	57	15	100
11	170	8	53	10	100	10	55	14	100
12	150	12	65	23	100	12	65	24	100
13	140	11	60	16	100	8	63	18	100
14	80	12	63	20	100	12	63	22	100
15	70	12	69	25	100	12	68	27	100
16	60	12	80	35	100	12	80	35	100
17	40	12	68	25	100	12	70	27	100
18	40	12	73	29	100	12	71	28	100
19	30	12	77	33	100	12	76	34	100
20[1]	80	10	60	17	0-100	10	60	17	0-100

Description of Numbered Locations

1. Room 123, antenna pointed to N.W. top corner of room
2. Room 198, North side
3. Hallway between Rooms 128 and 170
4. Room 128 West
5. Room 132
6. Hallway between Rooms 132 and 125
7. Room 125 West, antenna pointed to N-W corner
8. Room 126
9. Room 126 West, close to concrete wall.
10. Room 124 West
11. Room 124 near door, Antenna pointing toward door
12. Hallway near door to Room 122
13. Inside Room 122 (20 feet)
14. Hallway near Room 106 (restrooms). Note that there is a concrete wall to the West.
15. Room 150. Approximately 20 feet to the South of location 14, so the receiver is out of the "shadow" caused by the concrete wall.
16. Hallway outside entrance to Room 104.
17. Outside the building, West, 20 feet, antenna pointed to main glass door.
18. Same as loc. 17. Antenna pointed to steel entrance door to Room 102, i.e. more toward stationary antenna in Room 140.
19. Directly West of stationary antenna, just outside West external concrete wall.
20. Directly North of stationary antenna, just outside North external concrete wall.

Notes

1. By turning the cart, it was observed that signal quality was maintained over a 90 degree angle.

IV. CONCLUSIONS

The Persoft equipment used was found function satisfactorily and user-friendly. The equipment is suitable for incorporation into KSC networks for the purposes stated in the introduction. The maximum range specified for the equipment can be obtained only with line-of-sight contact between the two antennas. In practice, the range of the equipment is very much dependent on building materials for indoor use, and terrain for outdoor use. For any given application, careful consideration should be given to such factors as intervening masonry walls (for indoor use), intervening buildings, high terrain and trees (for outdoor use). This report incorporates general propagation data that should be helpful in evaluating each intended application. Based on the numerical data the propagation loss can be modeled as approximately 1 dB loss for every 7 feet. Concrete walls add approximately 5 dB loss. Outdoor propagation loss can be modeled as approximately 1 dB per 150 feet. The reception is improved in clear areas that are far away from buildings etc.

REFERENCES

[1] Ketchersid, John, "Bridges go wireless," *Network World*, June 12, 1995, p. 53.

[2] Anonymous, *Intersect Remote Bridge User Manual*, 1993, Madison, Wisc. : Persoft Inc.

[3] Stallings,William, *Data and Computer Communications, 4th ed.*, New York: MacMillan, 1994, pp. 488-495.

[4] Anderson, Jorgen B., Rappaport, T.S.,and Yoshida, S., "Propagation Measurements and Models for Wireless Communications Channels," *IEEE Communications Magazine,* January 1995, pp. 42-49.

1995 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

SG-26
7749
P- 24
(24)

CRACK GROWTH BEHAVIOR OF AISI-4340 STEEL DURING ENVIRONMENTAL EXPOSURE

Dr. Lucille A. Giannuzzi
Assistant Professor
Department of Mechanical and Aerospace Engineering
University of Central Florida
Orlando, Florida

KSC Colleague - Rupert Lee
Material Science

Contract Number NASA-NGT-60002
Supplement 19

August 9, 1995

PAGE 250 INTENTIONALLY BLANK

ACKNOWLEDGMENTS

I would like to express my sincere appreciation to the 1995 NASA/ASEE Summer Faculty Fellowship Program at Kennedy Space Center for their sponsorship of this research project. In addition, Dr. Ray Hosler, Ms. Kari Stiles, and Mr. Gregg Buckingham also deserve recognition for their administrative efforts throughout the summer, and for keeping the program running smoothly.

I would like to thank all of the members of the Materials Section of the Materials and Chemical Analysis Branch at KSC for stimulating discussions and invaluable help along the way. In particular, I would like to thank my NASA colleague, Dr. Rupert Lee for his guidance throughout the project, Mr. Peter Marciniak, for his help and expertise on metallographic specimen preparation and for enlightening discussions, Mr. Johnny Gay for machining anything on a moments notice, and to Mr. Heri Soto, for a memorable tour of LC-39a.

Specials thanks to Mr. Roger MacDonald, UCF graduate student, for his assistance and diligence on this project.

ABSTRACT

AISI-4340 is observed to undergo stress corrosion cracking when subjected to a constant load during exposure to a 3.5% NaCl solution. Crack initiation, nucleation, and growth has been monitored as a function of time. Stepped regions consisting of fast and slow crack growth periods are shown to correspond to microstructural changes observed in the fracture surface of the steel. These regions of fast and slow crack rate variations with time show that the crack growth rates do not increase continuously with an increase in the stress intensity.

SUMMARY

Compact tension specimens have been used to assess the stress corrosion cracking behavior of a high strength AISI-4340 steel. The crack tip was wet with a 3.5% NaCl solution and pulled in tension at a constant load until failure. The crack mouth opening displacement, the crosshead displacement, and the test duration was continuously monitored. Due to the plane stress conditions which were present during testing, K_{Iscc} and K_{Ic} could not be determined from this study. The crack initiation, nucleation, and growth rates may appear either as a gradual transition as time progresses until failure, or may yield a stepped behavior as time progresses. When a stepped behavior is observed, crack growth rates fluctuate between fast and slow values as K increases, which is not consistent with prior observations where the log(crack growth rate) increases continuously as K increases. Preliminary microstructural analyses of specimen fracture surfaces shows evidence of changing crack growth morphology in regions which are believed to correspond to the stepped crack growth behavior.

TABLE OF CONTENTS

LIST OF FIGURES

ABBREVIATIONS AND ACRONYMS LIST

a	crack length
B	C(T) specimen thickness
C(T)	compact tension specimen
CMO	crack mouth opening
E	Young's modulus
E'	effective Young's modulus
F	geometric constant
F'	geometric constant
HCL	hydrogen chloride
HRC	hardness Rockwell C-scale
K	stress intensity
K_{Ic}	critical stress intensity
K_{Iscc}	critical stress intensity for the onset of SCC
LEFM	linear elastic fracture mechanics
NaCl	sodium chloride
SCC	stress corrosion cracking
SEM	scanning electron microscope (microscopy)
UTS	ultimate tensile strength
W	length of a C(T) specimen from the load axis
YS	yield strength
α	the ration of a/W
ν	Poison's ratio
π	pi (3.14....)
σ	applied stress
σ_y	yield stress

I

INTRODUCTION

Environmental cracking or stress corrosion cracking (SCC) of a material is the result of a synergistic effect of exposure to a corrosive environment while under stress (i.e., a tensile stress). The combination of the environment and a tensile stress acting on a material is more deleterious than either the effect of the environment, or the effect of stress, acting alone. Thus, brittle catastrophic failure may occur in structural components in the presence of a corrosive environment even if the applied stresses are well below the yield strength of the material. The threat of SCC failures at NASA-KSC is a major concern for the shuttle and ground support materials due to the corrosive environment that these structural materials are subjected to. For example, the shuttle launch pads are located only 2500 feet (762 m) from the Atlantic Ocean, and a shuttle launch spews HCl as a by-product of the solid rocket booster combustion process. The abundance of chloride ions allows for the tremendous potential for corrosion failures on and around the launch facilities. Hence, it is very important to understand SCC mechanisms, so that materials selection and structural design can be incorporated into this harsh environment.

Structural materials are seldom loaded in a static condition. In addition, environmental conditions usually fluctuate. That is, structural materials are usually exposed to both cyclic loading conditions as well as cyclic environmental wet and dry periods. The ultimate goal of this project is to evaluate the effect of cyclic loads and exposures on AISI 4340 steel. However, the incubation and nucleation of crack growth during static loading and constant environmental exposure must first be understood.

The relative influences of electrochemical and mechanical factors in the corrosion and SCC damage of a material is shown in the crack growth vs. time curve in figure 1(ASM 1987). There exists an incubation and nucleation time period for crack growth which is primarily driven by electrochemical driving forces. Then, as time progresses, the driving force for crack growth leans toward the mechanical factor in the system. There is no accepted model for this particular crack growth phenomenon. That is, either a gradual transition of crack growth may occur until specimen failure as in figure 1, or, repeated succession of short steps of initiation and growth may occur until specimen failure.

The path of cracking in high strength steels is intergranular along prior austenite grain boundaries (Phelps and Loginow, Davis, and Dean and Copson). Mixed inter- and transgranular cracking has been observed in 4340M and 4330 steel (Davis et al.). In AISI 4340 steel, stress corrosion fracture surfaces were characterized as surface nucleated, intergranular fractures with secondary cracking or deep crevices (Phillips et al.). Transmission electron microscopy on AISI 4340 exposed to a 3% NaCl solution has shown that epsilon carbide precipitates were found to be preferentially attacked (Tiner and Gilpin). The attack was attributed to microstrains an/or depletion of alloying elements from the are adjacent to the carbide.

Linear-elastic fracture mechanics (LEFM) has been used to characterize the mechanical component of the driving force of SCC. The stress intensity, K, is the ability of a material to resist crack growth at an applied stress (i.e., the intensity of stress at the crack tip) and is given by:

$$K = F\sigma\sqrt{\pi a}$$

K = stress intensity (units of MPa $\sqrt{m}$)
a = crack length
σ = applied stress
F = f(geometry, a/W)

For a compact tension C(T) specimen, whose schematic diagram is shown in figure 2:

$$K = F' \frac{P}{W\sqrt{B}}$$

where:

B = specimen thickness
W = length of specimen from the load axis
F' = f(geometry, a/W) and $F \neq F'$

and

$$F' = \frac{(2+\alpha)}{(1-\alpha)3/2}(0.886 + 4.64\alpha - 13.32\alpha^2 + 14.72\alpha^3 - 5.6\alpha^4) \text{ where } \alpha = a/W$$

In order for LEFM to be applicable:

$$a, (W-a), h \geq \frac{4}{\pi}\left(\frac{K_{Ic}}{\sigma_y}\right)^2$$

where h = distance from the crack center line to the specimen edge, K_{Ic} is the critical fracture toughness, and σ_y is the yield strength.

For plane strain conditions:

$$B, a, (W-a), h \geq 2.5 \left(\frac{K_{Ic}}{\sigma_y}\right)^2$$

If plane strain conditions do not exist (i.e., plane stress conditions exist), K_{Ic} will be overestimated, because K_{Ic} increases as thickness, B, decreases. Therefore, the critical stress intensity factor should not be determined under plane stress conditions. The crack mouth opening displacement (CMO) is related to the material properties and crack growth as follows:

$$\frac{(CMO)}{P} = \left(\frac{1}{E'B}\right) q(a/W)$$

where q(a/W) = f(geometry, a/W), and E' is the effective Young's modulus (=E for plane stress, = $E(1-\nu^2)$ for plane strain). The crack growth rate is then related to the stress intensity as shown in figure 3 (ASM 1987). In stage I, the crack growth begins at a stress intensity = K_{Iscc} and the crack growth rate increases as K increases. In stage II, the crack growth rate is constant and independent of K and depends on the material/environment interaction. In stage III, the crack growth rate increases as K increases and approaches K_{Ic}.

There are several methods available for evaluating the stress corrosion susceptibility of a material. Each method, of course, has advantages and disadvantages. The method chosen in the present study consisted of constant load, K increasing, conditions. The advantage of these conditions is that they best mimic the loading conditions of structural members. The disadvantage to these parameters is the potentially long duration of each test.

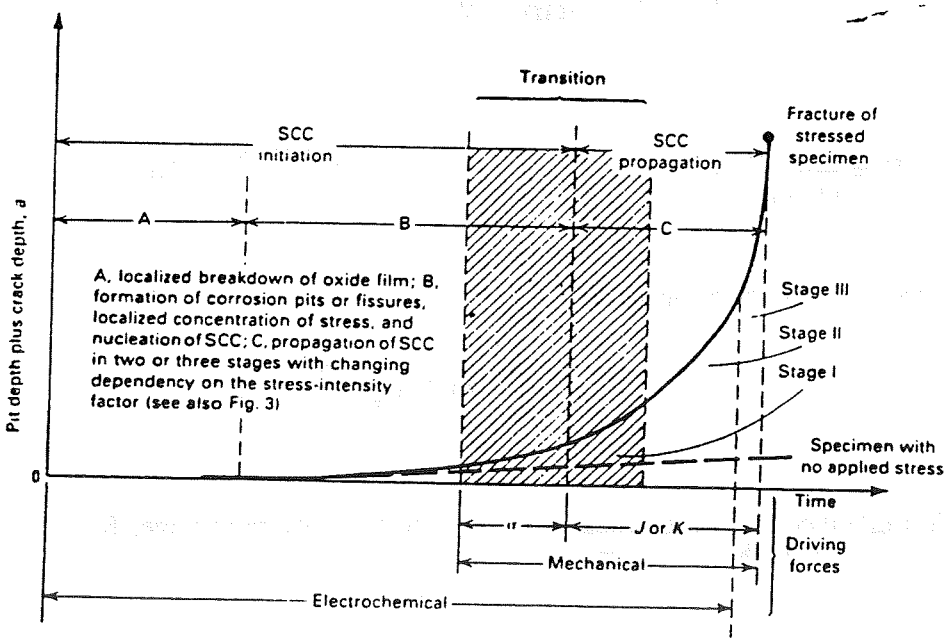


Figure 1. Schematic diagram of SCC crack growth behavior (from ASM 1987).

Compact Tension Specimen

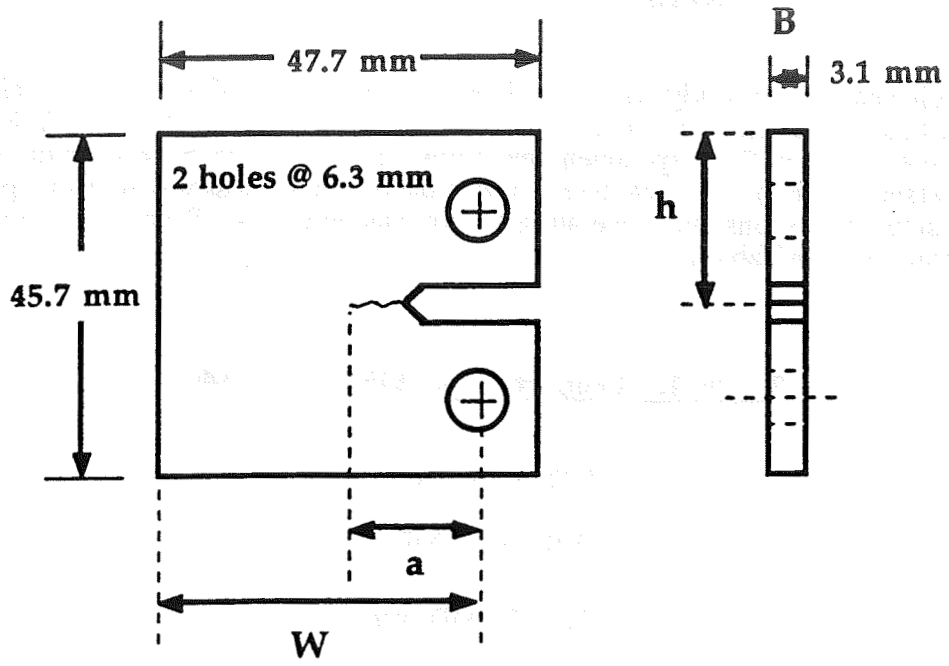


Figure 2. Schematic diagram of the C(T) specimen.

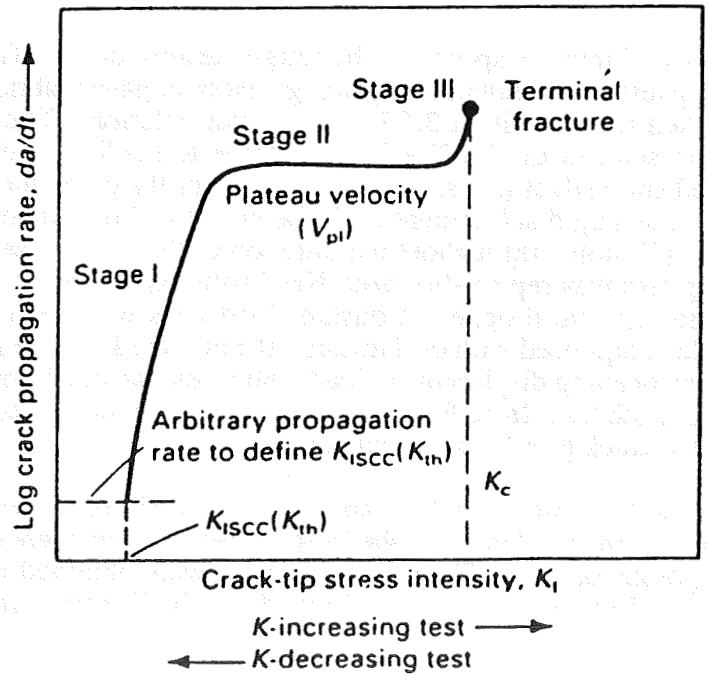


Figure 3. Schematic diagram of crack growth rates as a function of stress intensity (from ASM 1987).

II

EXPERIMENTAL PROCEDURES

The material chosen for this study was AISI-SAE 4340 steel plate of 3.1 mm in thickness. The steel was heat-treated to achieve a Rockwell hardness value of HRC = 42.8 ± 0.7. At least twenty-nine (29) compact tension C(T) specimens were machined from the plate according to ASTM standard E399 (see figure 2). It should be noted that the plate thickness was such that plane stress, and not plane strain conditions, were operating during tension testing. Several material properties of AISI-4340 are listed in Table 1.

Table 1. Properties of AISI-SAE 4340

UTS ~ 1500 MPa

YS ~ 1365 MPa

K_{Ic} ~ 75 MPa $\sqrt{m}$

K_{Iscc} ~ 30 MPa $\sqrt{m}$

Precracks were incorporated into the specimens by fatigue testing on an MTS Hydraulic Tester. Stress corrosion cracking tests were initiated by placing a piece of gauze fabric into the C(T) notch. The gauze fabric was then soaked with a 3.5% NaCl water solution. The C(T) specimen was loaded after a 0.5 hr pre-soak in the NaCl solution. Prior to loading, a clip strain gauge was fastened to the machined knife edges (not shown in figure 2) of the crack mouth of the specimen, so that the crack mouth opening displacement could be monitored throughout the test. The wet gauze remained in the C(T) notch throughout the duration of the test and served as the source of corrosive media. The gauze was replenished with NaCl solution as needed by squirting solution into the notch area using an eye dropper. Constant load tests were performed on a SATEC Universal Testing Machine equipped with the Nuvision II Software Package for automation of the testing. The crack mouth opening displacement, load, and crosshead displacement were monitored as a function of time, for each test. In addition, each test was video monitored and recorded so that an independent method of crack growth rates could be calculated.

The 4340 steel was polished for microstructural analysis using traditional metallographic polishing operations, followed by etching in Nital. Polished and etched surfaces were examined in a LECO metallograph. Macrographs of the C(T) fracture surfaces were obtained using a SONY CCD camera and video printer. Fractography was performed in a JEOL scanning electron microscope (SEM).

III

RESULTS AND DISCUSSION

Light optical micrographs of the microstructure of the 4340 steel are shown in figure 4. As evident from the micrographs, the microstructure of the steel consists primarily of martensite. The region in figure 4a is parallel to the direction of the crack plane, figure 4b shows the microstructure from the plane perpendicular to the crack direction and parallel to the C(T) thickness, and figure 4c shows the microstructure from the front surface of the C(T) specimen. Note that the bar on the micrographs corresponds to 25 μm.

Table 2 summarizes the SCC tests which were performed and lists the specimen identification number, the applied load, in kN, the total change in crack mouth opening displacement after loading (Δ CMO), in (m), and the length of time of each test at load in days, hours, minutes, and seconds. Test specimen 29 was the only specimen listed which did not fail in the time shown. Note that the time to failure does not correspond to the applied load, but depends on the initial applied stress intensity which varied with the initial crack length for each specimen (not listed in table 2).

Table 2. Summary of SCC Test Results

Specimen	Applied Load (kN)	Δ CMO (m)	Time (d:hh:mm:ss)
11	7.8	2.9×10^{-7}	0:01:01:52
13	9.8	1.6×10^{-5}	0:01:06:20
28	5.9	2.5×10^{-4}	0:05:39:36
14	5.9	9.1×10^{-5}	0:06:36:00
07	8.9	3.2×10^{-4}	1:01:00:11
29	5.9	6.4×10^{-7}	1:22:35:24 + (no failure)

The CMO vs. time for each specimen listed in Table 2 are plotted and shown in figures 5-10. Note that only specimen 14 shows a smooth transition from the initiation and nucleation stage of a crack, to crack growth, and subsequent failure. All of the other tests show one or more stepped crack growth arrest periods prior to fast fracture of the specimen. Note that SCC testing of specimen 29 was interrupted and therefore, the test did not result in specimen failure. Due to time constraints, the data could not be presented in the traditional manner which plots the log of the crack growth rate as a function of instantaneous stress intensity (as in figure 3). However, it is evident that as

the crack grows, the stress intensity, K, increases. Thus, while K is constantly increasing throughout the duration of a test, the velocity of the crack may increase or decrease in a stepped fashion as evident by the crack growth arrest periods in figures 5-10. Hence, the traditional stage I, stage II, and stage III portions of the log crack rate vs. K behavior during SCC is only observed for specimen 14 in this study.

The entire fracture surface from specimen 13 is shown in figure 11, and is typical of all of the SCC fracture surfaces which were observed. Extending from right to left, the fracture surface consists of the notch, the fatigue precrack region, the SCC region (darkest area), and the fast fracture region (largest and brightest area showing cleavage).

Fractography was performed on all of the SCC specimens and select SEM micrographs are shown below. The typical stress corrosion morphology which was evident in all of the specimens is shown in the SEM micrograph of figure 12a. Figure 12b shows a typical region which has undergone fast fracture. A low magnification SEM micrograph from specimen 28 is shown in figure 13. Note that the fatigue precrack region, the stress corrosion cracking region, and the region of fast fracture is contained within the frame of micrograph extending from left to right. The stress corrosion cracking microstructure can be divided into alternate regions of slow and fast crack growth which correspond to the CMO vs. time curve as shown by the arrows in figure 13. Figure 14 shows the surface of specimen 29 which did not fail during testing. Note the plastic zone associated with local deformation at the crack tip.

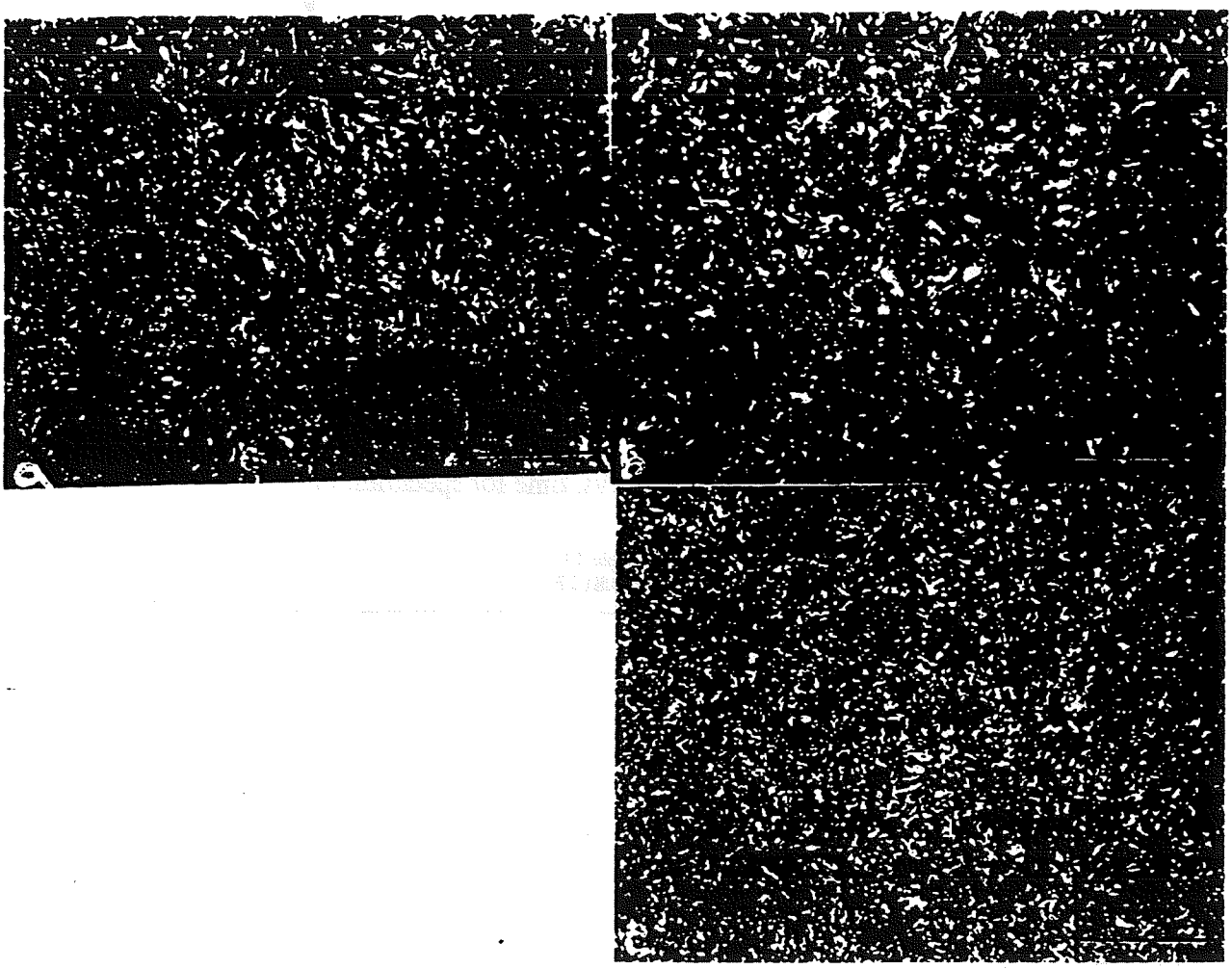


Figure 4. Polished and etched surfaces of the AISI-4340 C(T) specimen from a) parallel to the crack plane. b) perpendicular to the crack plane. c) the C(T) surface.

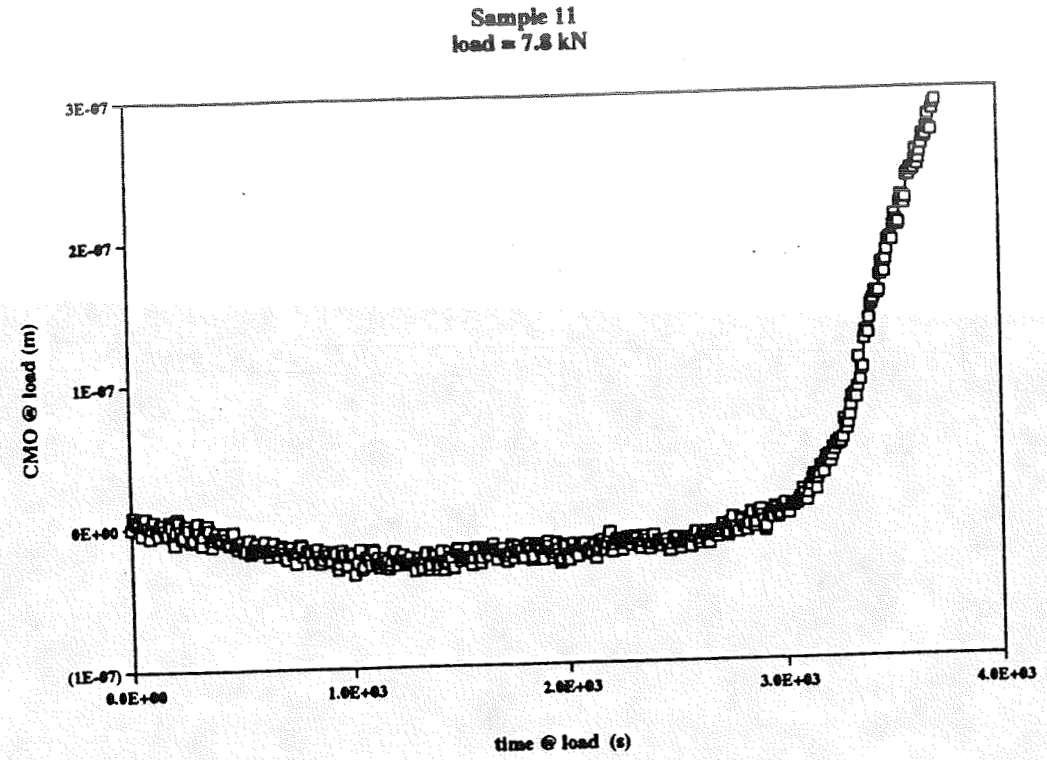


Figure 5. Crack Mouth Opening Displacement vs. time for specimen 11.

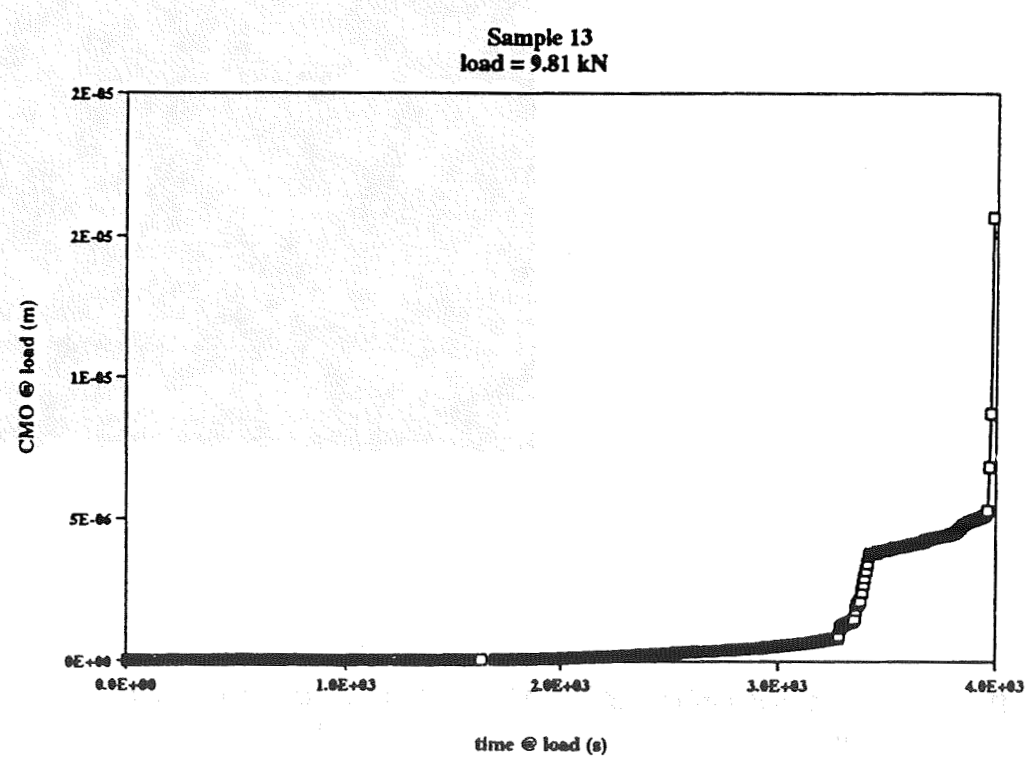


Figure 6. Crack Mouth Opening Displacement vs. time for specimen 13.

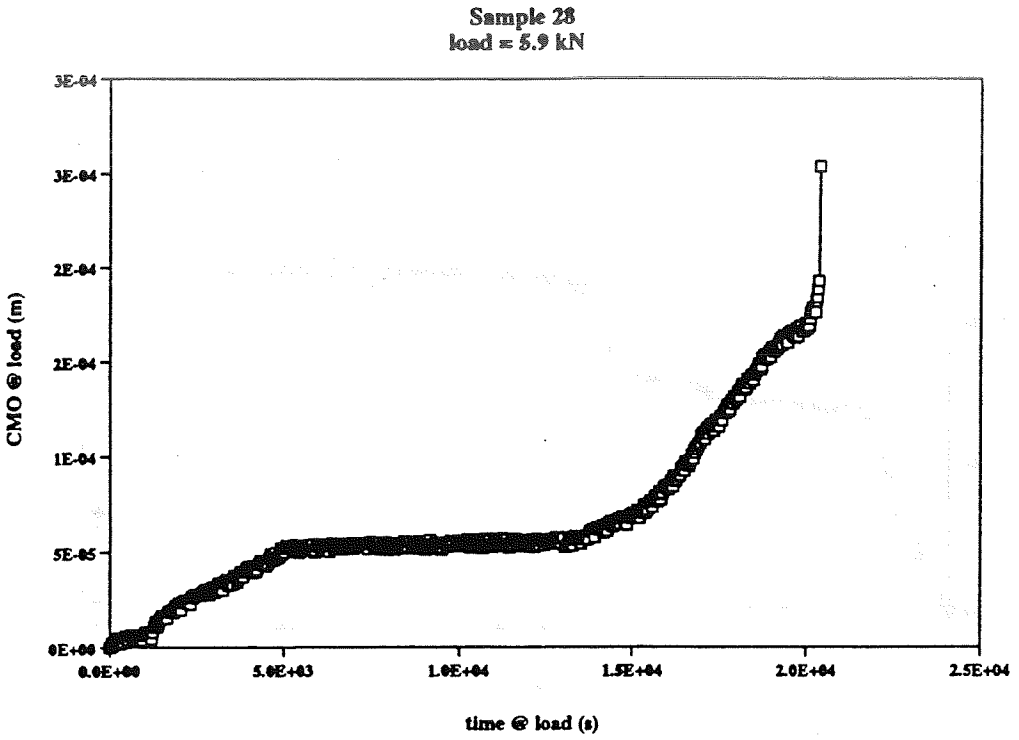


Figure 7. Crack Mouth Opening Displacement vs. time for specimen 28.

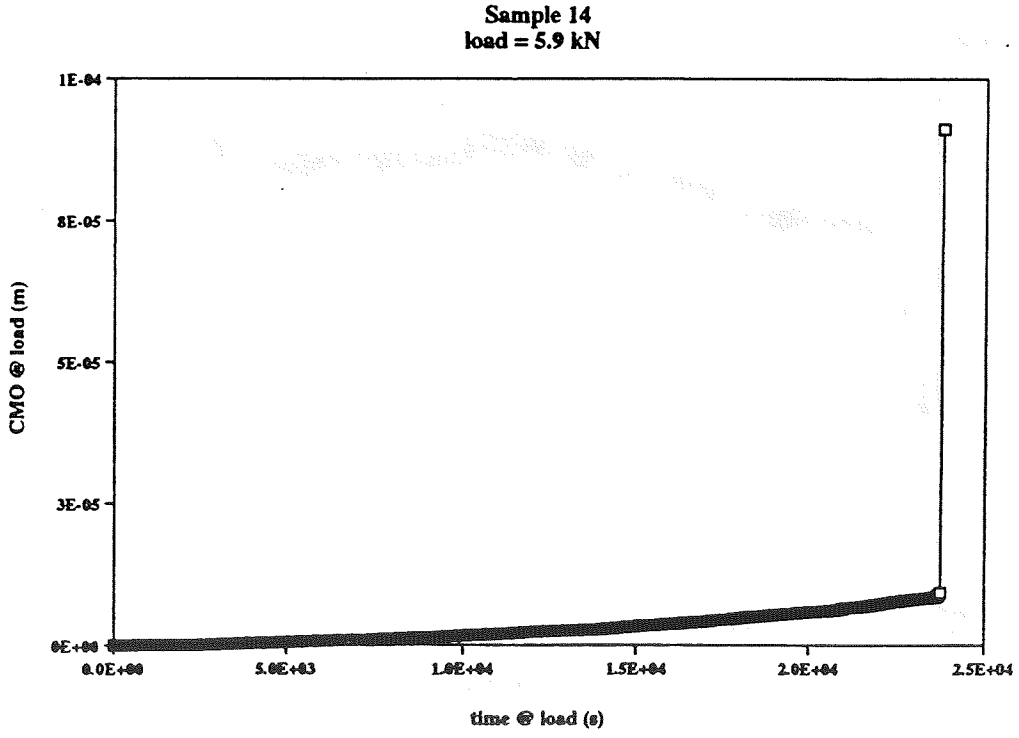


Figure 8. Crack Mouth Opening Displacement vs. time for specimen 14.

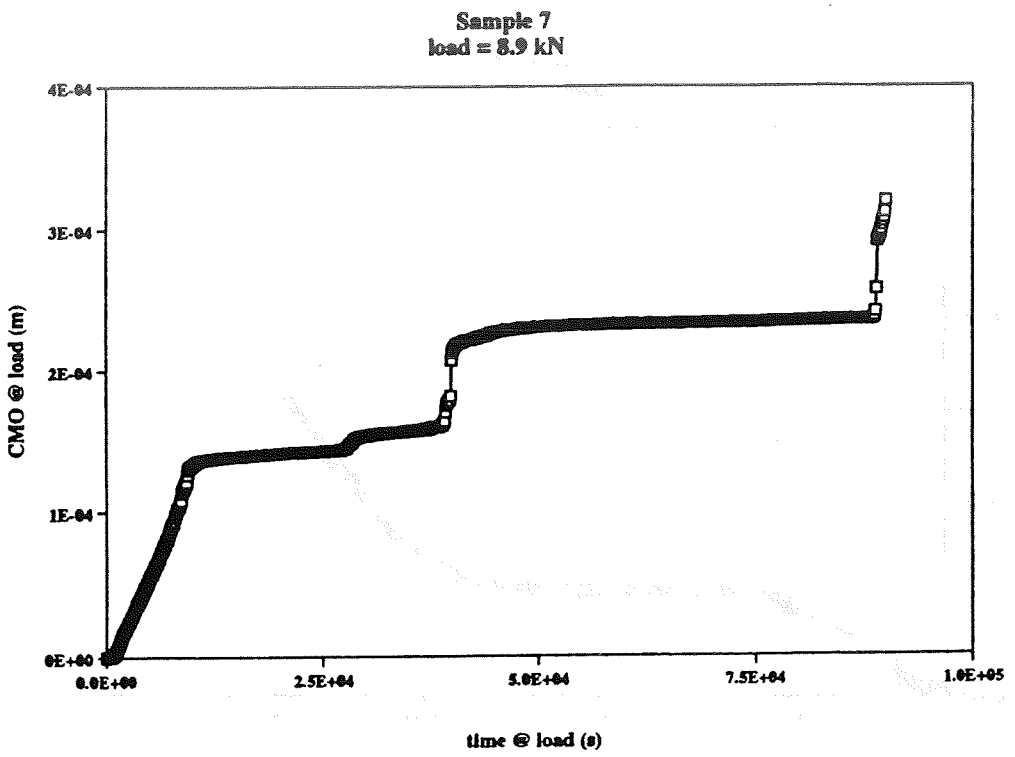


Figure 9. Crack Mouth Opening Displacement vs. time for specimen 7.

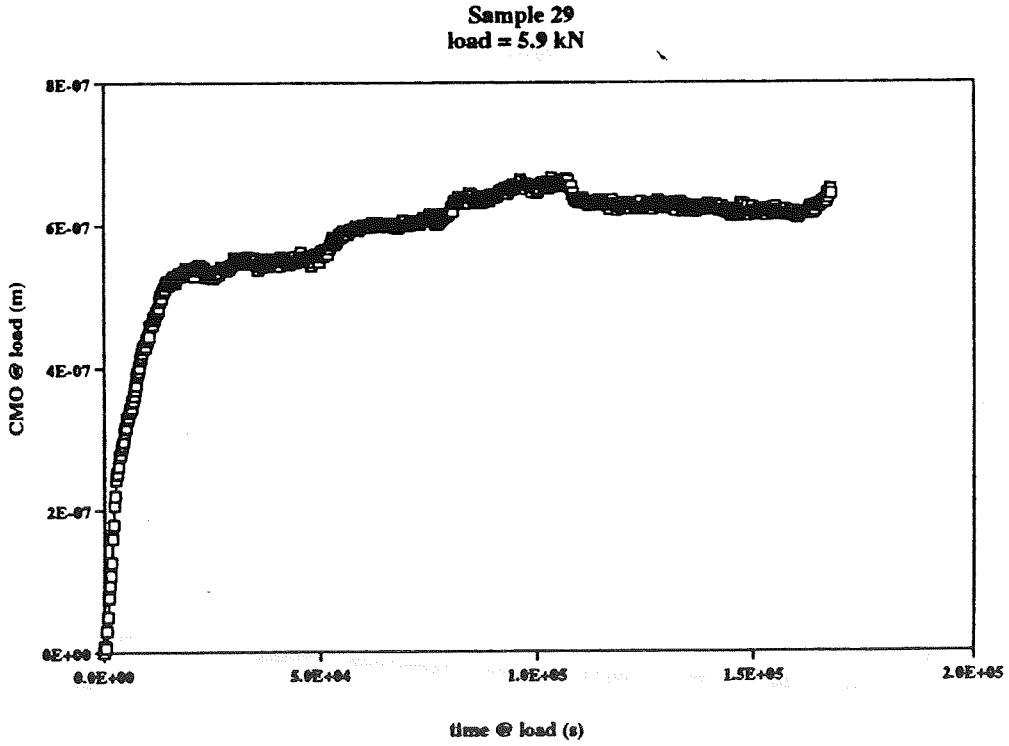


Figure 10. Crack Mouth Opening Displacement vs. time for specimen 29.

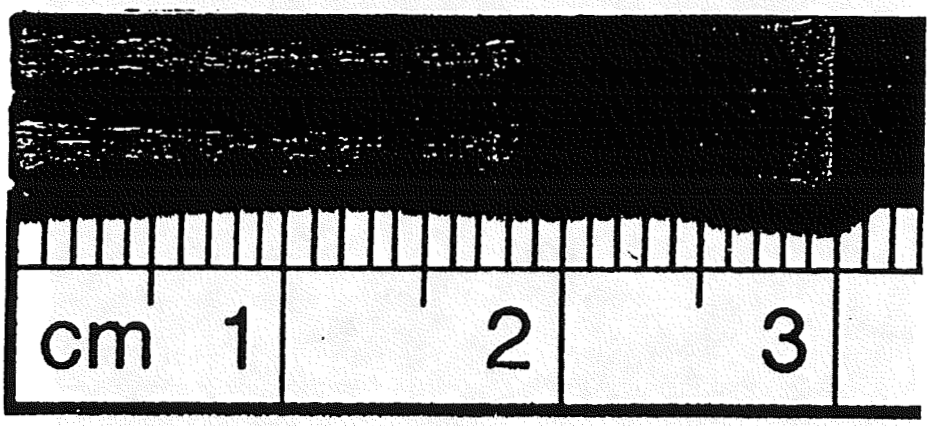


Figure 11. Fracture surface of specimen 13.

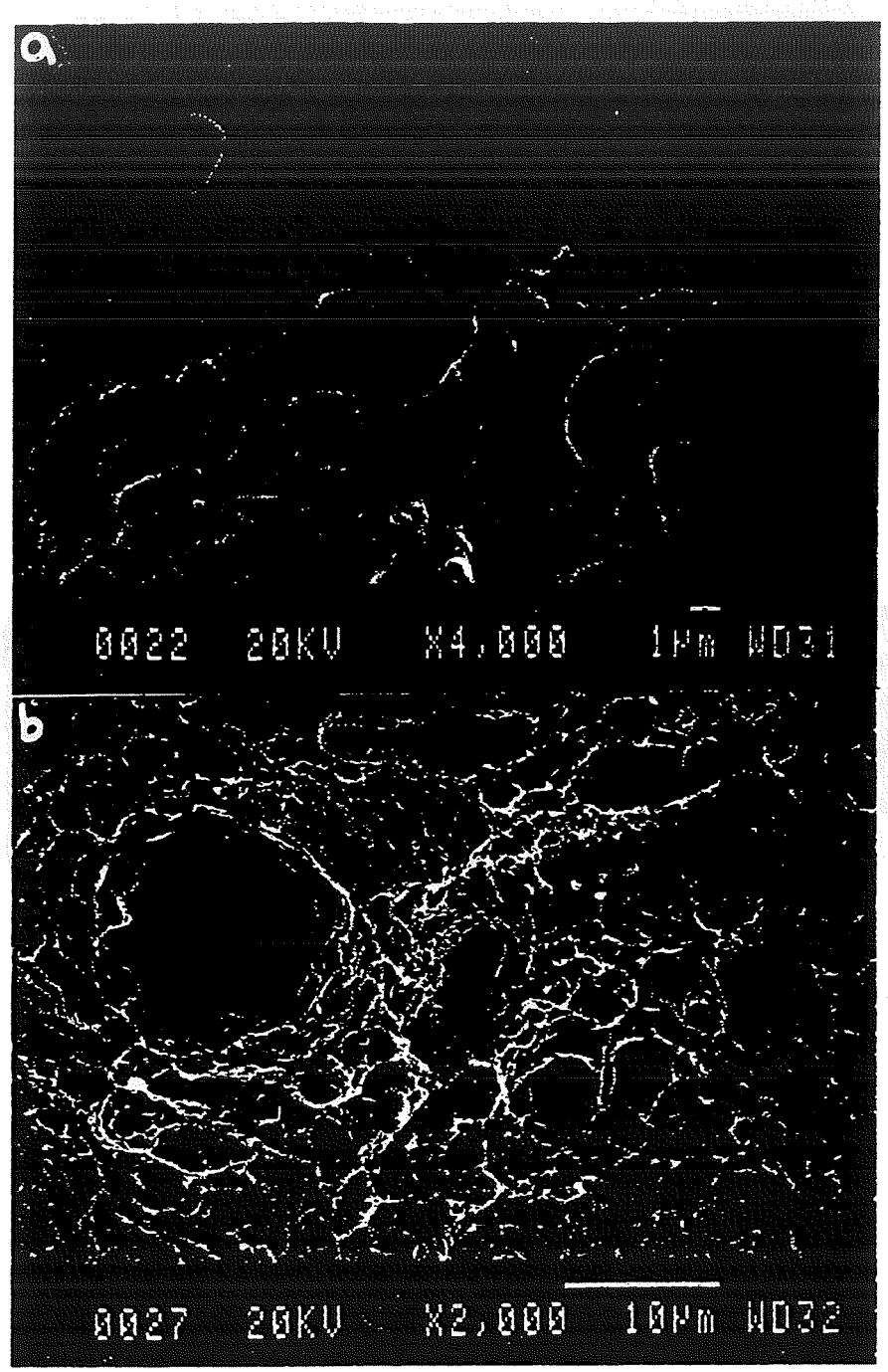


Figure 12. SEM micrograph of a typical SCC fracture surface from: a) the SCC region b) the fast fracture region.

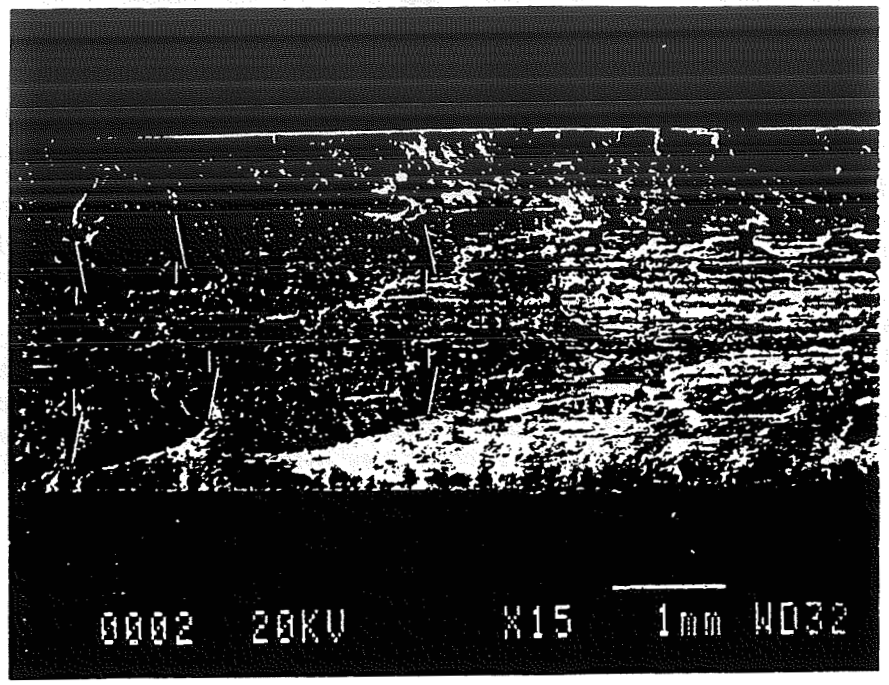


Figure 13. SEM micrograph of the fracture surface of specimen 28.

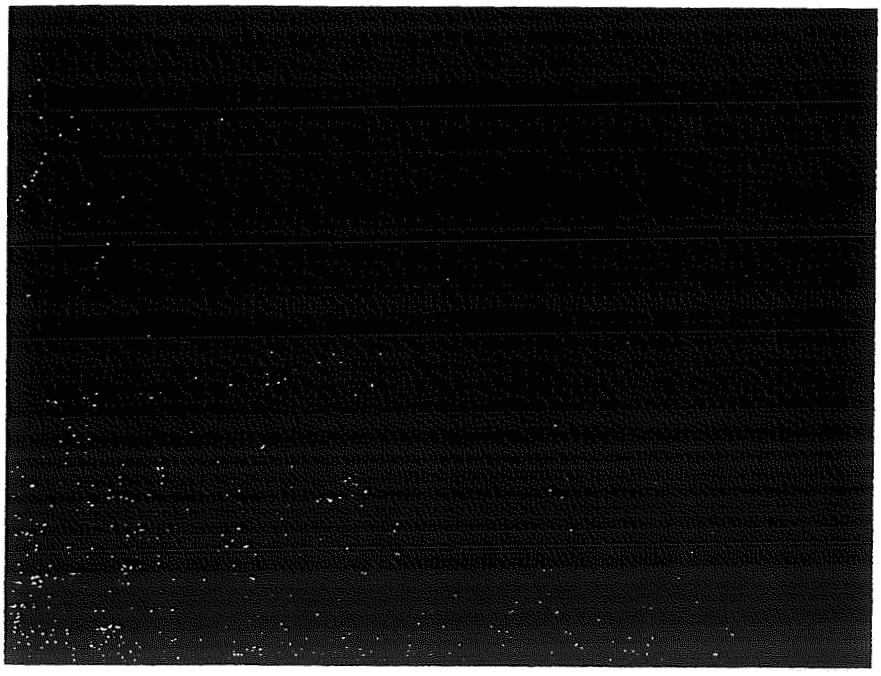


Figure 14. The plastic zone associated with local deformation at the crack tip during crack growth.

IV

CONCLUSIONS

SCC of AISI 4340 steel has been observed during constant loading conditions when exposed to 3.5% NaCl solution. Due to the plane stress conditions which were present during testing, K_{Iscc} and K_{Ic} could not be determined from this study. The crack initiation, nucleation, and growth rates may appear either as a gradual transition as time progresses until failure, or may yield a stepped behavior as time progresses. When a stepped behavior is observed, crack growth rates fluctuate between fast and slow values as K increases, which is not consistent with prior observations where the log(crack growth rate) increases continuously as K increases. Preliminary microstructural analyses of specimen fracture surfaces shows evidence of changing crack growth morphology in regions which are believed to correspond to the stepped crack growth behavior.

REFERENCES

ASM Metals Handbook, 9th ed. vol. 13, (1987), p. 246.

R.A. Davis, "Stress Corrosion Investigation of Two Low Alloy High Strength Steels," Corrosion, 19, 45t-54t, (1963).

R.A. Davis, G.A. Dreyer, and W.C. Gallaugher, "Stress Corrosion Cracking Study of Several High Strength Steels," Corrosion, 20, 93t-103t (1964).

S.W. Dean and H.R. Copson, "Stress Corrosion Behavior of Maraging Nickel Steels in Natural Environments," Corrosion, 21, 95-103 (1965).

E.H. Phelps and A.L. Loginow, "Stress Corrosion of Steels for Aircraft and Missiles," Corrosion, 16, 325t-335t, (1960).

N.A. Tiner and C.B. Gilpin, "Microprocesses in Stress Corrosion of Martensitic Steels," Corrosion, 22, 271-279 (1966).

1995 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

510-71
7750
p. 30
30

CAVITATION EFFECTS IN ULTRASONIC CLEANING BATHS

Dr. Barbara H. Glasscock
Associate Professor
Chemical and Materials Engineering Department
California State Polytechnic University, Pomona
Pomona, California

KSC Colleague - Gale Allen
Material Science

Contract Number NASA-NGT-60002
Supplement 19

August 18, 1995

PAGE 274 INTENTIONALLY BLANK

ACKNOWLEDGMENTS

I would like to express my sincere appreciation to the Office of Educational Affairs, NASA Headquarters, Washington, D.C. and the American Society for Engineering Education (ASEE) for their sponsorship of the 1995 NASA/ASEE Summer Faculty Fellowship Program at Kennedy Space Center. My summer research experience at KSC has been stimulating and rewarding and I know that I will carry with me to my home university many of the new ideas I have gained here.

The program was well organized and ran smoothly, thanks to the efforts of Dr. Ray Hosler and Ms. Kari Stiles, both of the University of Central Florida and to Gregg Buckingham of NASA/KSC.

A very warm thanks to all of the folks in the Materials and Chemistry Branch at KSC for their help enabling me to accomplish all that I did. Gale Allen, my NASA Colleague, was very supportive and often helped remove road blocks that appeared along the way. Bryan Feddersen, a co-op student from Georgia Tech, worked closely with me on this project, and was a valuable colleague. The SEM photos which were so revealing and intriguing were taken by Virginia Cummings and Stan Young. X-ray photoelectron spectroscopy is credited to Orlando Melendez. Laura Bales helped with the coupon acquisition and preparation. And many more thanks to the others at KSC with whom I interacted.

I am indebted to the folks at the Component Refurbishment and Chemical Analysis Laboratory at KSC operated by Wiltech of Florida. Bob LaVoie was the originator of the idea of using aluminum foil in a frame to measure the bath intensity and also conducted much of the coupon testing for us. Also thanks to Ken Fishell and Sam Kershner of Wiltech for their assistance and ideas.

I would also like to express my appreciation to my husband, John, for his support and willingness to relocate to Cocoa Beach for the summer so that I could pursue this work.

ABSTRACT

In this project, the effect of cavitation from aqueous ultrasonic cleaning on the surfaces of metal and non-metal sample coupons was studied. After twenty cleaning cycles, the mass loss from the aluminum coupons averaged 0.22 mg/cm^2 surface area and 0.014 mg/cm^2 for both stainless steel and titanium. The aluminum coupons showed visual evidence of minor cavitation erosion in regions of previously existing surface irregularities. The non-metal samples showed some periods of mass gain. These effects are believed to have minor impact on hardware being cleaned, but should be evaluated in the context of specific hardware requirements. Also the ultrasonic activity in the large cleaning baths was found to be unevenly distributed as measured by damage to sheets of aluminum foil. It is therefore recommended that items being cleaned in an ultrasonic bath be moved or conveyed during the cleaning to more evenly distribute the cavitation action provide more uniform cleaning.

SUMMARY

Aqueous ultrasonic cleaning has been proposed as a replacement at KSC for CFC solvents which deplete stratospheric ozone. In ultrasonics, cavitation bubbles are created by high frequency vibrations which can cause stresses at solid surfaces. The purpose of this study was to see if the effects from ultrasonic cavitation bubbles was severe enough to damage the surfaces of metals and non-metals being cleaned. If damage from ultrasonics was found, it would limit the applicability of this process in the intended use for cleaning of flight and ground support hardware at KSC.

Coupons of stainless steel, aluminum, titanium, Vespel™ and Viton™ were exposed to a full-scale ultrasonic cleaning process consisting of four ultrasonic baths. The first bath operated at 27 kHz and contained the degreaser Brulin™, the remaining baths all operated at 40 kHz. The second and fourth baths contained only deionized water and the third bath contained the surfactant Zonyl™. After twenty cleaning cycles, the mass losses from the metal were very low: 0.22 mg/cm^2 surface area for the aluminum coupons and 0.014 mg/cm^2 for both stainless steel and titanium. The aluminum coupons showed visual evidence of minor cavitation erosion in regions of previously existing surface irregularities. The non-metal samples showed some periods of mass gain.

The distribution of ultrasonic activity throughout the large cleaning baths was evaluated using sheets of aluminum foil spanning the entire widths of the baths. Aluminum foil is easily damaged by ultrasonics and the pattern of damage to the foil was used to map the areas of greater and lesser ultrasonic intensity. The sheets of aluminum foil were placed approximately 10" apart in the baths. The ultrasonic intensity in the baths was found to be far from uniform. The aluminum foil sheets in Bath A, operating at 27 kHz, showed a greater amount of damage than those in the other baths operating at 40 kHz. Within a given bath there was a great amount of variation on a single sheet as well as from adjacent sheets. These findings suggest that parts being cleaning in ultrasonic baths should be moved throughout the field to obtain the best cleaning efficiency. Evidence of the presence of standing wave patterns were observed on some of the aluminum foil sheets from two of the ultrasonic baths, B and CR, and the space between clusters of holes agreed with the theoretical value for the wavelength.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

I. INTRODUCTION

1.1 Background

Researchers at Kennedy Space Center and elsewhere have been working to develop replacement chemicals for chlorofluorocarbons (CFCs) because of the link between CFCs and the depletion of the protective layer of ozone in the stratosphere. For example, 1,1,2 trichloro - 2,2,1 trifluoroethane (CFC-113) was used at the Kennedy Space Center for precision cleaning of aerospace hardware. Many of the replacement cleaning systems that have been developed have focused on aqueous cleaning solutions which would be more environmentally acceptable.

One of the replacement cleaning systems includes the use of ultrasonics with aqueous solutions to enhance the cleaning. In ultrasonics, cavitation bubbles are created at the solid/liquid interface by high frequency vibrations. The stresses caused by the formation and collapse of the bubbles loosens debris from the surface, and the turbulence caused by the bubbles helps to transport the debris into the bulk solution. [Reference 1]

Ultrasonic cavitation can produce high temperatures and pressures that are very localized and are present for short durations of time. The stresses at the solid surface due to the cavitation are severe enough to potentially corrode the solid surface. A previous study [Reference 2] investigated the effect of pure water and ultrasonics on the surfaces of small precision parts for cleaning verification. Using Scanning Electron Microscopy and weight loss measurements, the study concluded that the corrosion effects and mass losses were insignificant for ultrasonic exposure periods of up to 2 hours.

An ultrasonic cleaning system proposed for cleaning of flight hardware at the Component Refurbishment and Chemical Analysis Laboratory (CRCA) at the Kennedy Space Center includes the use of surfactants, Brulin™ or Zonyl™, which reduce the surface tension of the water, enhancing the cleaning of small cavities. This system involving surfactants has not been evaluated as to its possible effects on the surfaces of the parts being cleaned.

1.2 Purpose

The purpose of this study was to evaluate the effects of cavitation from ultrasonic cleaning on the surfaces of metals and non-metal coupons. It will be necessary to assure minimal effects from ultrasonic cleaning before the process can be approved for cleaning of flight hardware. The purpose of the second part of this study was to evaluate the uniformity of the ultrasonics in the baths. The motivation for the uniformity study was to ensure that coupons being studied would be placed in areas of high ultrasonic activity.

II. MATERIALS AND METHODS

2.1 Ultrasonic Baths

The ultrasonic cleaning baths used in this study are used for cleaning of aerospace hardware at the Component Refurbishment and Chemical Analysis Laboratory (CRCA) located at the Kennedy Space Center. The baths are part of a cleaning process that has been developed as a replacement for a vapor degreasing process that used CFC-113.

Four large ultrasonic baths, as described in Table 2-1, are operated in series. The ultrasonic transducers are in the bottoms of the tanks. Parts to be cleaned are placed into baskets which rest on support racks, approximately eight to ten inches from the transducers. The baskets are moved from bath to bath after each ultrasonic cycle. The solutions within each tank are circulated through filtration systems to remove particulates. During actual cleaning operation, the parts may be placed into one of various chemical baths between immersion in Bath B and Bath C.

Table 2-1. Ultrasonic Bath Descriptions

Bath Number	Ultrasonic Power	Work Area (inches)	Volume (gal)	Solution	Temperature
A	9600 watts @ 27 kHz	65 x 36	260	Brulin 815 GD	150°F
B	7200 watts @ 40 kHz	54 x 36	205	D.I. water	150°F
C	4800 watts @ 40 kHz	48 x 28	160	DuPont Zonyl FSN	150°F
CR	3600 watts @ 40 kHz	36 x 28	120	D.I. water	150°F

2.2 Metal and Non-Metal Coupons

Two different shapes of coupons were used in this test: large rectangular coupons to simulate larger parts being cleaning, and small disks to simulate smaller parts. The coupon specifications are summarized in Table 2-2.

2.3 Coupon Preparation

The large coupons were previously engraved with identification numbers. The small disks were labeled with a handheld engraving tool, except the Viton™ disks which were identified with notches in the edges. Because this test was only concerned with cavitation effects and not cleaning efficiency, it was desirable to start with clean coupons in order to use weight changes as a measure of material loss. All of the metal coupons were cleaned in advance of the study by rinsing first with CFC-113 and then acetone.

Table 2-2. Test Coupon Specifications

Large Coupons (rectangular)	Material Type	Rockwell B Hardness (HRB) average	Rockwell B Hardness (HRB) range	Approx. Dimensions (cm)
Aluminum	3003	20.7	14.6 - 23.8	3.7 x 6.5
Stainless Steel	304	85.1	84.3 - 85.7	4.0 x 6.3
Titanium	-	108.1*	107.2 - 108.6*	3.7 x 6.3
Small Coupons (disks)				Diameter (cm)
Aluminum	T6-6061	55.6	53.0 - 57.1	1.8
Stainless Steel	316L	97.2	95.7 - 99.4	1.8
		Shore durometer "A" scale		
Vespel™	-	97	96 -98	2.2
Viton™	-	79	76 - 82	1.8

* Note: Values of HRB over 100 are usually converted to Rockwell C scale.

2.4 Coupon Arrangement

Most of the large coupons were bolted to the washing baskets so that they would remain stationary, simulating the cleaning of large parts. The large coupons were fabricated with a hole (approximately 1 cm diameter) at one end of the rectangle, used for the bolting. Teflon washers were used to eliminate abrasion. Six large coupons were bolted to each of the baskets, designated α and β. One coupon of each material type (aluminum, stainless steel, and titanium) was positioned vertical and one horizontal in each basket. All of the small coupons plus one large coupon of each material were placed in a mesh basket with a lid that allowed some movement during the ultrasonics. Thus, a total of five large coupons of each material were tested and four small disks of each material. The different placements and orientations of the coupons were chosen to see if any of these factors played a role in the cavitation effects. Figure 2-1 shows the placement of the baskets within the ultrasonic baths. These locations were chosen because they represented areas of relatively high ultrasonic activity as described in Section 3.2.1. Care was taken to always place the coupons in the same locations on the baskets, which was placed in the same locations in the baths. (The designations A6, A5, etc. shown in Figure 2-1 were used for the aluminum foil tests and will be discussed in Section 2.7.)

2.5 Coupon Test Procedure

The coupons were exposed to a number of ultrasonic cleaning cycles with evaluations performed after 1, 10, and 20 complete cycles. Twenty cycles was chosen to be a reasonable upper bound for the number of times that a piece of hardware would be returned during its lifetime to the CRCA for cleaning. A complete cycle refers to 3 minutes of ultrasonification at 150°F (65.5°C) in each of the four cleaning baths (A, B, C, CR). Before evaluation of the coupons, the coupons were further rinsed with deionized water and allowed to dry.

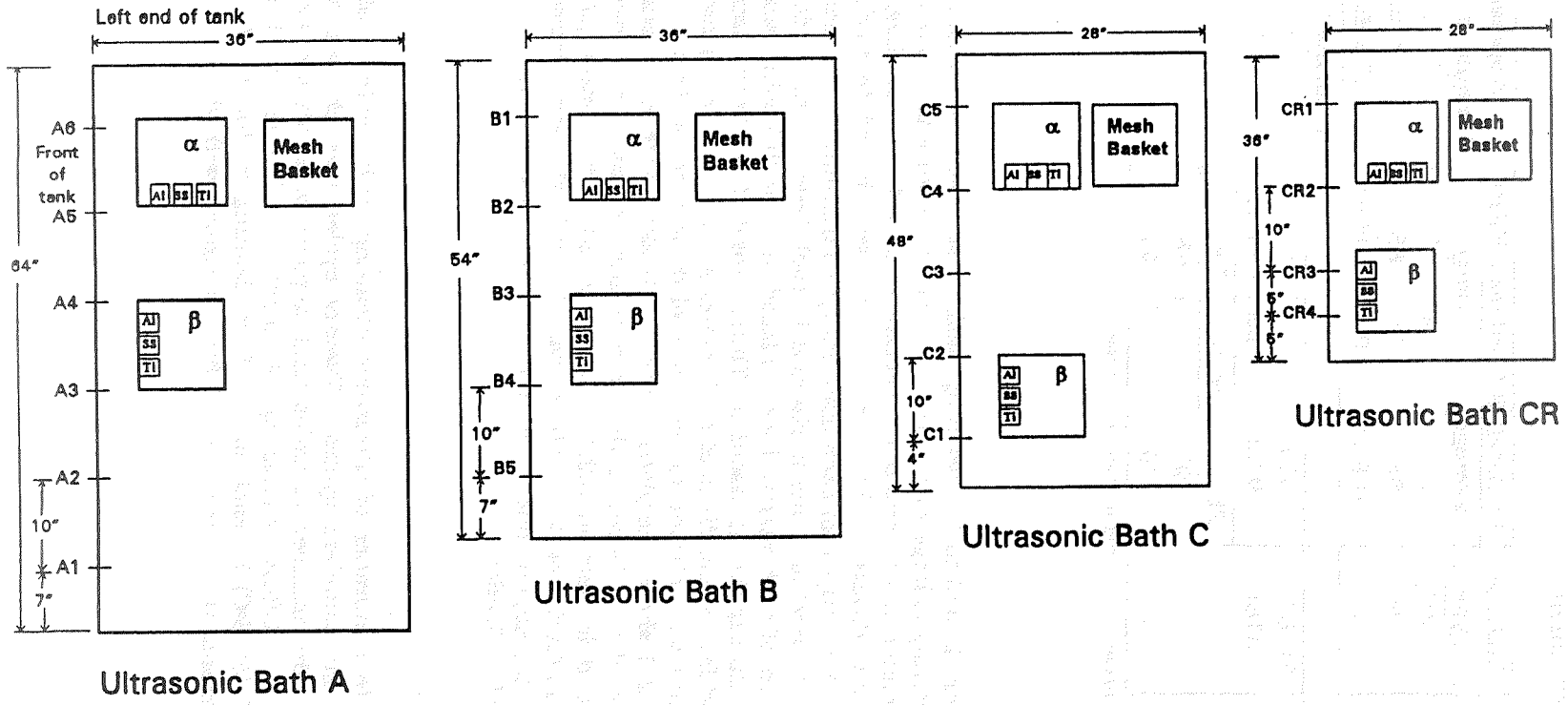


Figure 2-1. Placement of test coupon baskets in the ultrasonic baths (locations A6, A5,... were used for aluminum foil tests, see Section 2.7.)

2.6 Coupon Evaluation

Before the first ultrasonic cleaning cycle and after 1, 10, and 20 complete cycles, the coupons were weighed to evaluate for mass loss. One coupon of each material type was also analyzed using Scanning Electron Microscopy (SEM) and X-ray Photoelectron Spectroscopy (XPS). To facilitate viewing exactly the same microstructure on the metals using SEM, a microhardness mark was placed on the surface as a reference, each mark being 200 to 300 microns in length, depending on the hardness of the metal.

2.7 Aluminum Foil Test Procedure

In ultrasonic applications, strips of aluminum foil are sometimes used to evaluate ultrasonic intensity because of the ease with which the foil erodes. In this study rather than using small strips, large sheets of aluminum foil were used to obtain a full cross-sectional cut of the bath to develop a map of the activity in the bath. The sheets of aluminum foil were held taut in a frame such as those used to hold window screens. The sheets of aluminum foil were placed in the baths, one at a time, approximately 10 inches apart, for a duration of three minutes of ultrasonification. Figure 2-1 shows the placement of the foil sheets by the marks A6, A5, etc. The lower numbers (A1, B1, C1, CR1) indicate the inlet of the recirculated liquid.

2.8 Aluminum Foil Evaluation

After exposure to the ultrasonic baths, the aluminum foil sheets were examined visually for cavitation effects. Each 4" x 4" square of a sheet was given a rating as to the damage to the foil and then a map was compiled from the data. In addition, several localized areas from selected sheets were examined using the SEM.

III. RESULTS AND DISCUSSION

3.1 Coupon Tests

3.1.1 Material Loss

All of the coupons were weighed initially and after 1, 10, and 20 cleaning cycles. This data is presented in Table 3-1, along with the basket identification (α, β, or Mesh) and the orientation of the coupon (Vertical or Horizontal). The aluminum coupons lost between 0.05% and 0.11% mass after 20 cycles, depending on the alloy and coupon dimensions. If the data is expressed as (mass loss/surface area) then the data for both coupon types is very close with an average of 0.22 mg/cm^2. The surface area used is an apparent surface area using the area of just the two faces of the coupon. No difference in mass loss was observed for vertical versus horizontal orientation and only a slight difference between basket location.

The stainless steel and titanium coupons showed very low mass loss rates with average losses after 20 cycles of 0.0014% and 0.0019% respectively. When expressed as (mass loss/surface area) the stainless steel and titanium coupons lost an average of 0.014 mg/cm^2.

The behavior of the non-metal coupons was more complex. The Vespel™ gained mass after 10 cycles, presumably absorbing liquid, and then lost some of its mass between 10 and 20 cycles, but still ended above its original weight. The Viton™ coupons lost weight after 1 cycle but then appear to have absorbed liquid, gaining weight after 10 and 20 cycles. (Note that Table 3-1 reports mass loss and thus negative values are actually mass gain.)

The mass loss data is summarized in Figures 3-1 and 3-2 where the mass loss/surface area is presented graphically as a function of the number of cycles. The data for the two different types of aluminum have been combined and the data for the stainless steels and titanium have been combined.

3.1.2 Scanning Electron Micrographs

Scanning Electron Micrographs (SEMs) were taken of one coupon of each material type before exposure to the ultrasonics (initial) and after 1, 10, and 20 cycles. Very little change was observed in all cases. Figures 3-3 to 3-7 present the SEM's of the coupons in the initial conditions and after exposure to 20 ultrasonic cleaning cycles. In all of these photos, the diamond-shaped mark in the center is the microhardness mark that was used for reference enabling examination of exactly the same microstructure each time.

In Figures 3-3 and 3-4, some of the inclusions of the aluminum alloys appear more evident after exposure to the ultrasonics. The large stainless steel coupon in Figure 3-5 shows no damage from the ultrasonics with the only effect being perhaps that the surface appears cleaner. The small stainless steel coupon, shown in Figure 3-6, appears to have small trails of very fine dents which may have been caused by vibrations in the mesh basket. The titanium coupon showed no damage after 20 cycles as seen in Figure 3-7.

Table 3-1. Material loss due to ultrasonic cleaning.

sample #	Basket	orientation	Surface Area (cm^2)	weights (g)				mass loss (mg)			% mass loss			mass loss/S.A. (mg/cm^2)		
				initial	1 cycl	10 cycl	20 cycl	1 cycl	10 cycl	20 cycl	1 cycl	10 cycl	20 cycl	1 cycl	10 cycl	20 cycl
Aluminum - large coupons																
Al 5-01*	α	H	46.08	18.7562	18.7554	18.7499	18.7453	0.8	6.3	10.9	0.0043	0.0336	0.0581	0.017	0.137	0.237
Al 5-03	α	V	46.23	18.6673	18.6663	18.6611	18.6563	1.0	6.2	11.0	0.0054	0.0332	0.0589	0.022	0.134	0.238
Al 5-04	β	H	46.48	18.8794	18.8789	18.8742	18.8693	0.5	5.2	10.1	0.0026	0.0275	0.0535	0.011	0.112	0.217
Al 5-05	β	V	46.53	18.8520	18.8516	18.8469	18.8419	0.4	5.1	10.1	0.0021	0.0271	0.0536	0.009	0.110	0.217
Al 5-02	M	H	46.50	18.8749	18.8743	18.8687	18.8635	0.6	6.2	11.4	0.0032	0.0328	0.0604	0.013	0.133	0.245
						Al-large avg ->		0.7	5.8	10.7	0.0035	0.0308	0.0569	0.014	0.125	0.231
Aluminum - small coupons																
Al-A3*	M	H	4.98	1.0173	1.0173	1.0167	1.0163	0.0	0.6	1.0	0.0000	0.0590	0.0983	0.000	0.120	0.201
Al-A6	M	H	4.98	1.0185	1.0187	1.0180	1.0174	-0.2	0.5	1.1	-0.0196	0.0491	0.1080	-0.040	0.100	0.221
Al-12	M	H	4.98	1.0177	1.0176	1.0172	1.0166	0.1	0.5	1.1	0.0098	0.0491	0.1081	0.020	0.100	0.221
Al-17	M	H	4.98	1.0148	1.0148	1.0142	1.0139	0.0	0.6	0.9	0.0000	0.0591	0.0887	0.000	0.120	0.181
						Al-small avg ->		0.0	0.5	1.0	-0.0025	0.0541	0.1008	-0.005	0.110	0.206
Stainless Steel - large coupons																
SS 3-57*	α	H	47.41	60.0458	60.0461	60.0455	60.0453	-0.3	0.3	0.5	-0.0005	0.0005	0.0008	-0.006	0.006	0.011
SS 3-74	α	V	49.14	62.0991	62.0991	62.0991	62.0989	0.0	0.0	0.2	0.0000	0.0000	0.0003	0.000	0.000	0.004
SS 3-99	β	H	49.24	62.3486	62.3483	62.3481	62.3476	0.3	0.5	1.0	0.0005	0.0008	0.0016	0.006	0.010	0.020
SS 3-47	β	V	47.64	60.2772	60.2773	60.2770	60.2767	-0.1	0.2	0.5	-0.0002	0.0003	0.0008	-0.002	0.004	0.010
SS 3-92	M	H	48.89	61.8543	61.8537	61.8530	61.8524	0.6	1.3	1.9	0.0010	0.0021	0.0031	0.012	0.027	0.039
						SS-large avg ->		0.1	0.5	0.8	0.0002	0.0007	0.0013	0.002	0.009	0.017
Stainless Steel - small coupons																
SS-A1*	M	H	4.86	3.3842	3.3842	3.3841	3.3841	0.0	0.1	0.1	0.0000	0.0030	0.0030	0.000	0.021	0.021
SS-A6	M	H	4.98	3.4340	3.4339	3.4339	3.4339	0.1	0.1	0.1	0.0029	0.0029	0.0029	0.020	0.020	0.020
SS-A7	M	H	4.98	3.2874	3.2874	3.2874	3.2874	0.0	0.0	0.0	0.0000	0.0000	0.0000	0.000	0.000	0.000
SS-A9	M	H	4.86	3.2053	3.2053	3.2053	3.2053	0.0	0.0	0.0	0.0000	0.0000	0.0000	0.000	0.000	0.000
						SS-small avg ->		0.0	0.0	0.0	0.0007	0.0015	0.0015	0.005	0.010	0.010

Table 3-1. Material loss due to ultrasonic cleaning (cont).

sample #	Basket	orientation	Surface Area (cm^2)	weights (g)				mass loss (mg)			% mass loss			mass loss/S.A. (mg/cm2)		
				initial	1 cycl	10 cycl	20 cycl	1 cycl	10 cycl	20 cycl	1 cycl	10 cycl	20 cycl	1 cycl	10 cycl	20 cycl
Titanium - large coupons																
TI-29*	α	H	45.30	34.1122	34.1119	34.1114	34.1114	0.3	0.8	0.8	0.0009	0.0023	0.0023	0.007	0.018	0.018
TI-20	α	V	45.65	34.2335	34.2335	34.2331	34.2330	0.0	0.4	0.5	0.0000	0.0012	0.0015	0.000	0.009	0.011
TI-21	β	H	45.21	34.0092	34.0092	34.0089	34.0085	0.0	0.3	0.7	0.0000	0.0009	0.0021	0.000	0.007	0.015
TI-13	β	V	46.00	34.2849	34.2849	34.2847	34.2846	0.0	0.2	0.3	0.0000	0.0006	0.0009	0.000	0.004	0.007
TI-15	M	H	46.32	34.5995	34.5992	34.5990	34.5986	0.3	0.5	0.9	0.0009	0.0014	0.0026	0.006	0.011	0.019
						TI avg ->		0.1	0.4	0.6	0.0003	0.0013	0.0019	0.003	0.010	0.014
Vespel - small coupons																
Vespel A1*	M	H	7.60	0.9109	0.9108	0.9126	0.9120	0.1	-1.7	-1.1	0.0110	-0.1866	-0.1208	0.013	-0.224	-0.145
Vespel A5	M	H	7.60	0.9055	0.9054	0.9071	0.9068	0.1	-1.6	-1.3	0.0110	-0.1767	-0.1436	0.013	-0.211	-0.171
Vespel A6	M	H	7.60	0.9078	0.9077	0.9095	0.9090	0.1	-1.7	-1.2	0.0110	-0.1873	-0.1322	0.013	-0.224	-0.158
Vespel A13	M	H	7.60	0.9116	0.9115	0.9133	0.9129	0.1	-1.7	-1.3	0.0110	-0.1865	-0.1426	0.013	-0.224	-0.171
						Vespel avg ->		0.1	-1.7	-1.2	0.0110	-0.1843	-0.1348	0.013	-0.220	-0.161
Viton - small coupons																
Viton 100*	M	H	4.86	0.7488	0.7481	0.7484	0.7484	0.7	0.4	0.4	0.0935	0.0534	0.0534	0.144	0.082	0.082
Viton 200	M	H	4.86	0.6805	0.6798	0.6800	0.6801	0.7	0.5	0.4	0.1029	0.0735	0.0588	0.144	0.103	0.082
Viton 300	M	H	4.98	0.7533	0.7526	0.7527	0.7528	0.7	0.6	0.5	0.0929	0.0796	0.0664	0.141	0.120	0.100
Viton 500	M	H	4.86	0.7289	0.7282	0.7283	0.7285	0.7	0.6	0.4	0.0960	0.0823	0.0549	0.144	0.123	0.082
						Viton avg ->		0.7	0.5	0.4	0.0963	0.0722	0.0584	0.143	0.107	0.087
* denotes SEM analysis done																

Figure 3-1. Mass loss/surface area for metal coupons

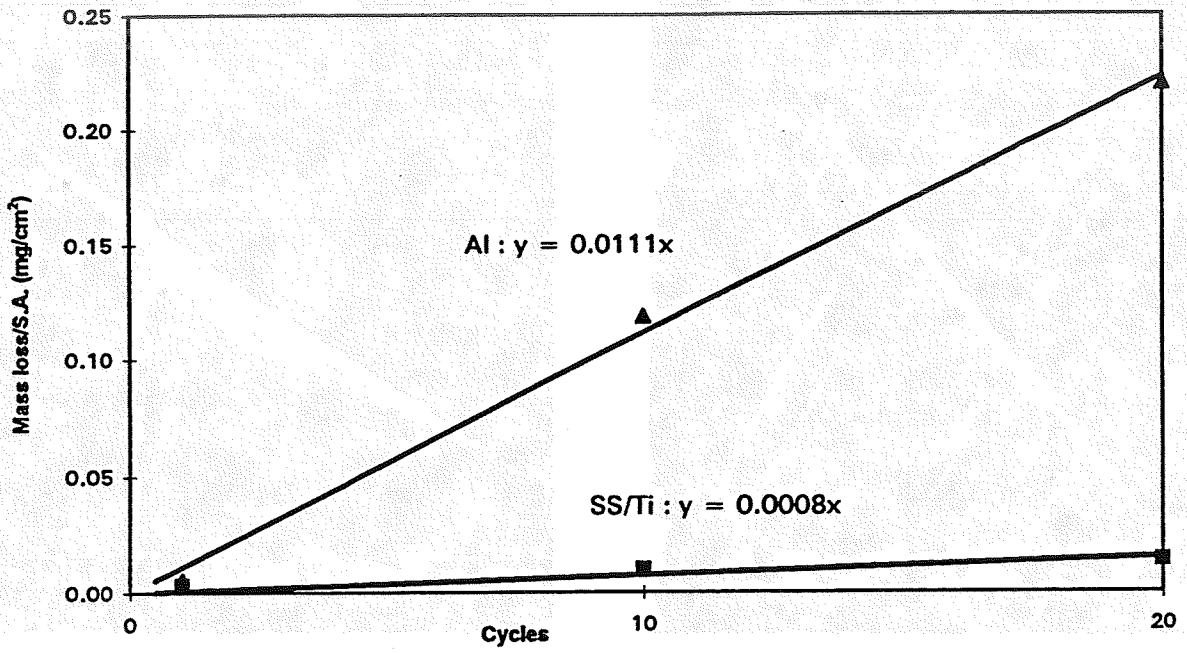
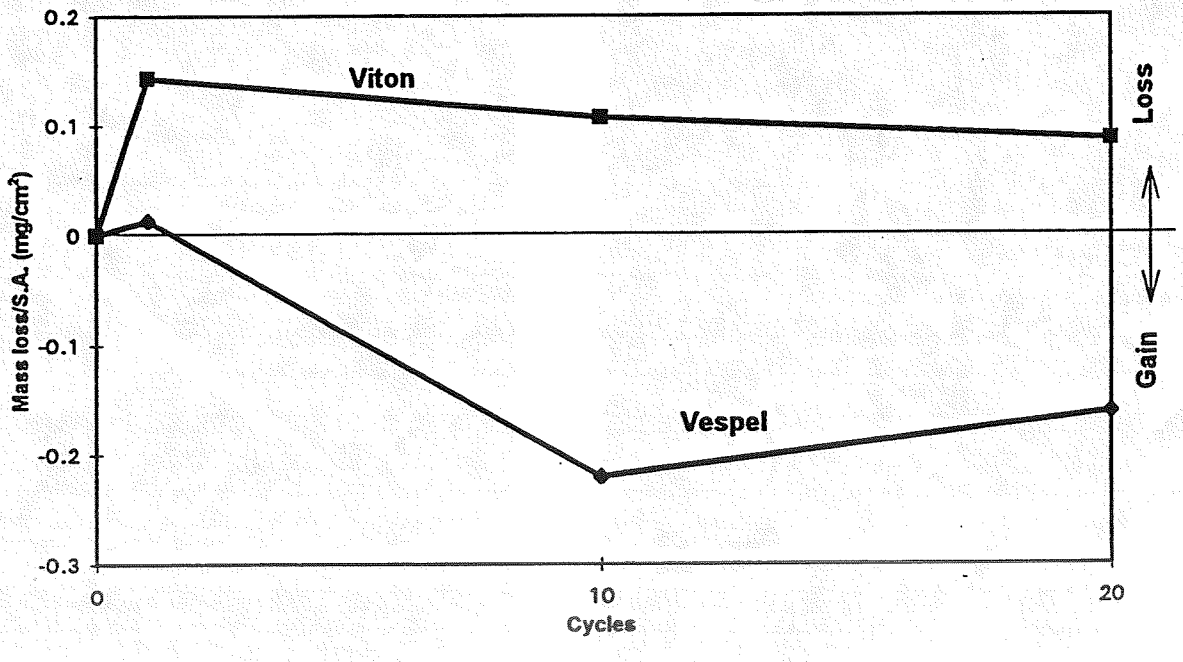


Figure 3-2. Mass loss/S.A. vs. cycles for non-metals



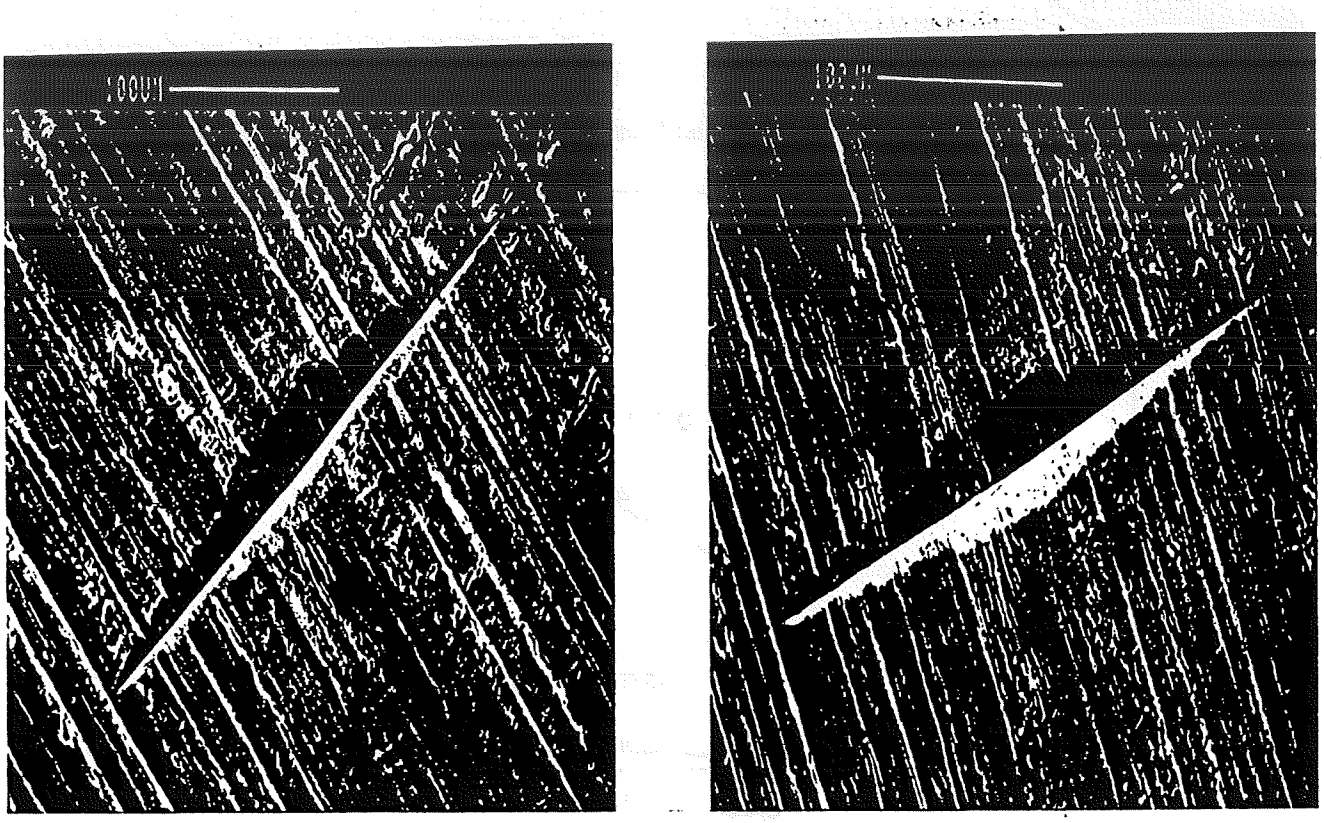


Figure 3-3. SEM of Aluminum (sample AL5-01) in its initial condition (left) and after 20 ultrasonic cleaning cycles (right). Microhardness mark in the center is for reference.

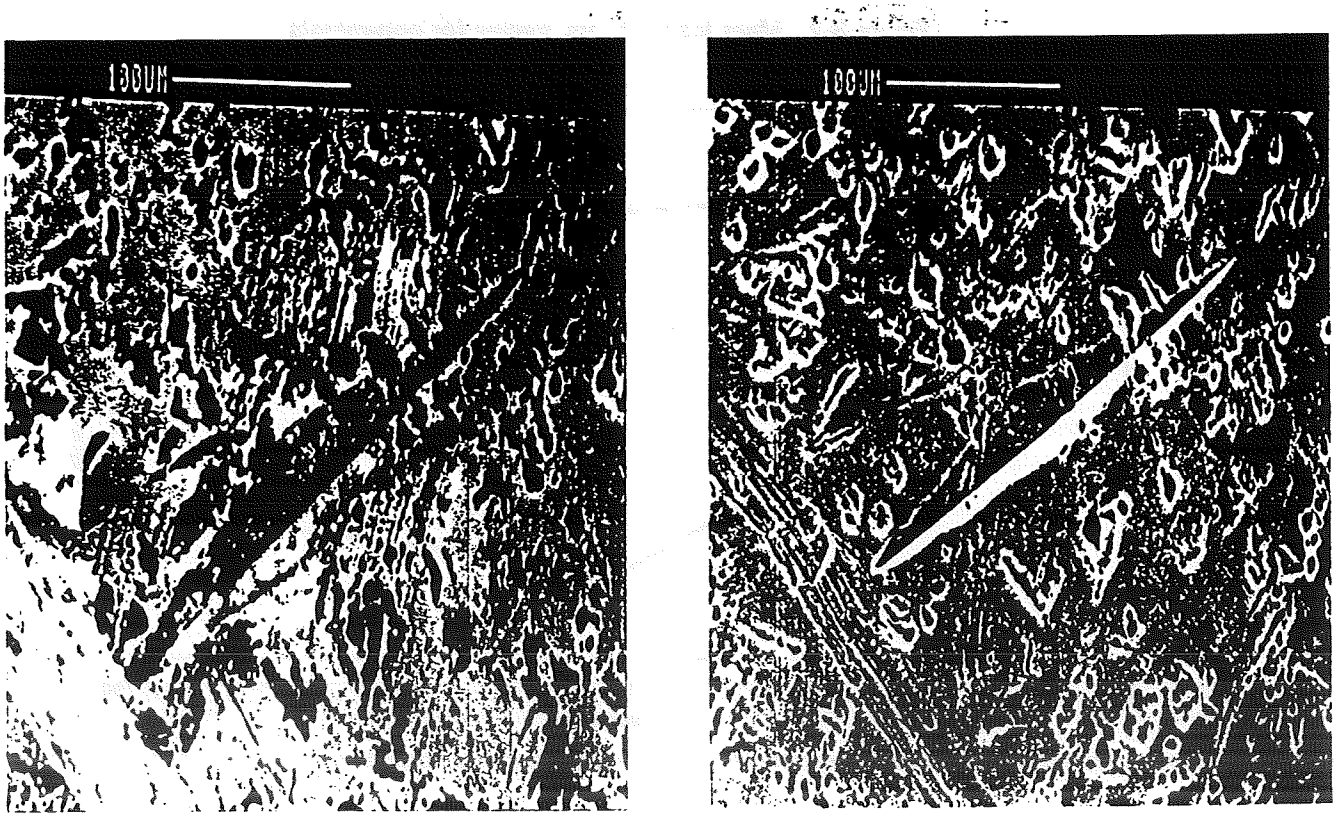


Figure 3-4. SEM of Aluminum (sample AL-A3) in its initial condition (left) and after 20 ultrasonic cleaning cycles (right).

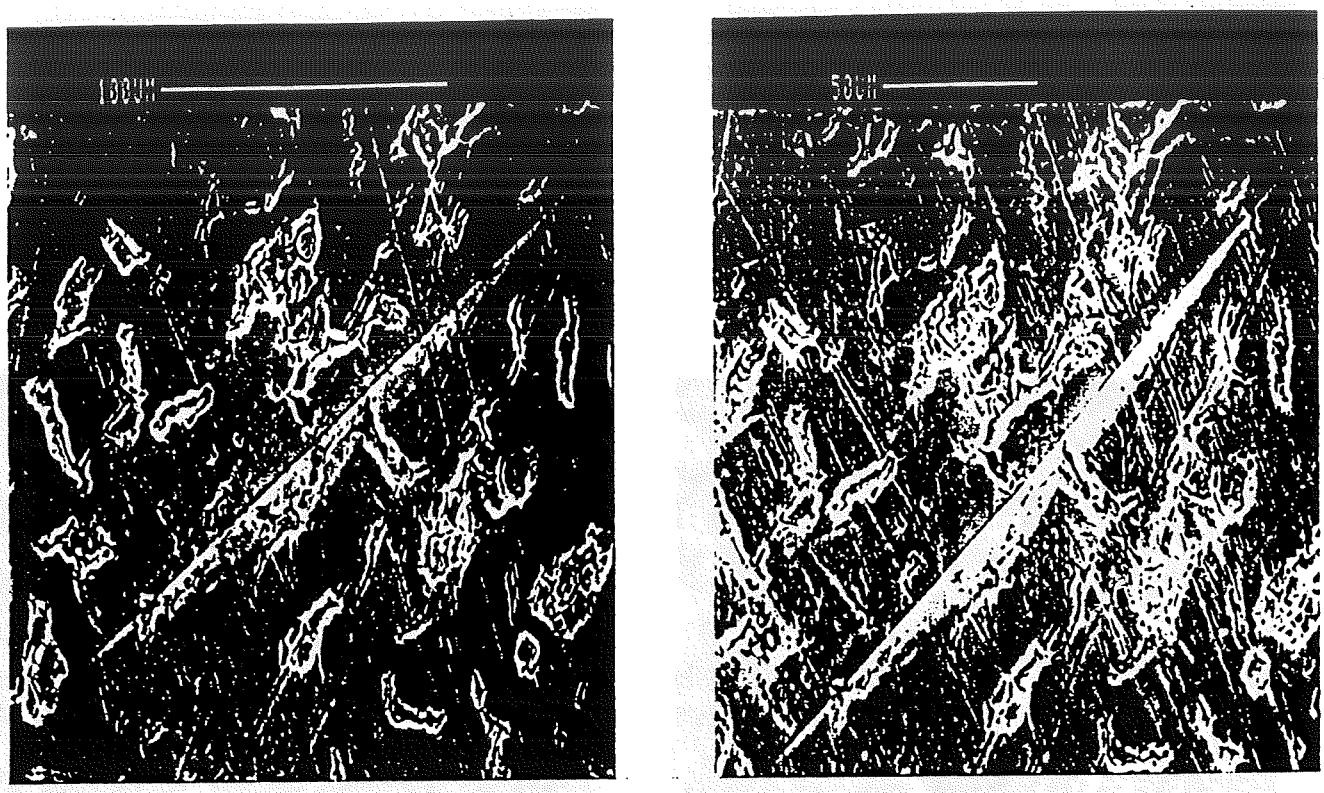


Figure 3-5. SEM of Stainless Steel (sample SS3-57) in its initial condition (left) and after 20 ultrasonic cleaning cycles (right).

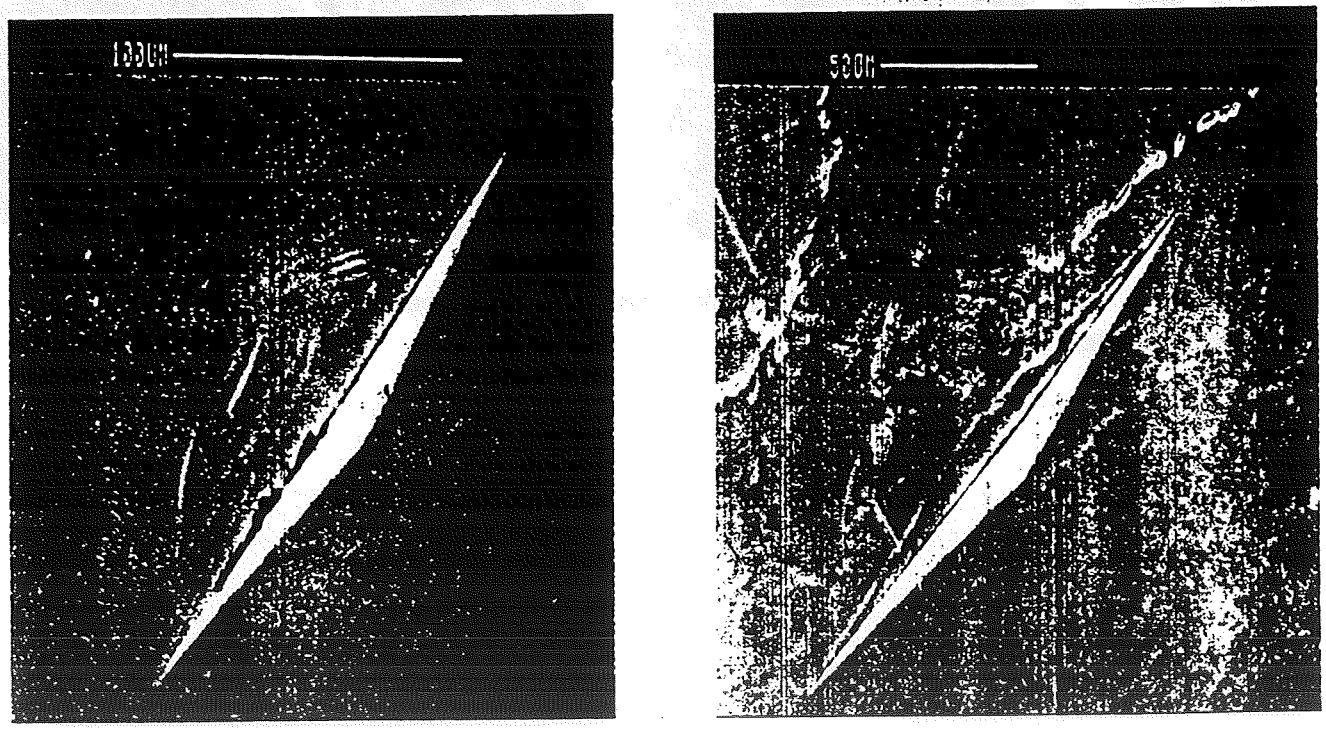


Figure 3-6. SEM of Stainless Steel (sample SS-A1) in its initial condition (left) and after 20 ultrasonic cleaning cycles (right).

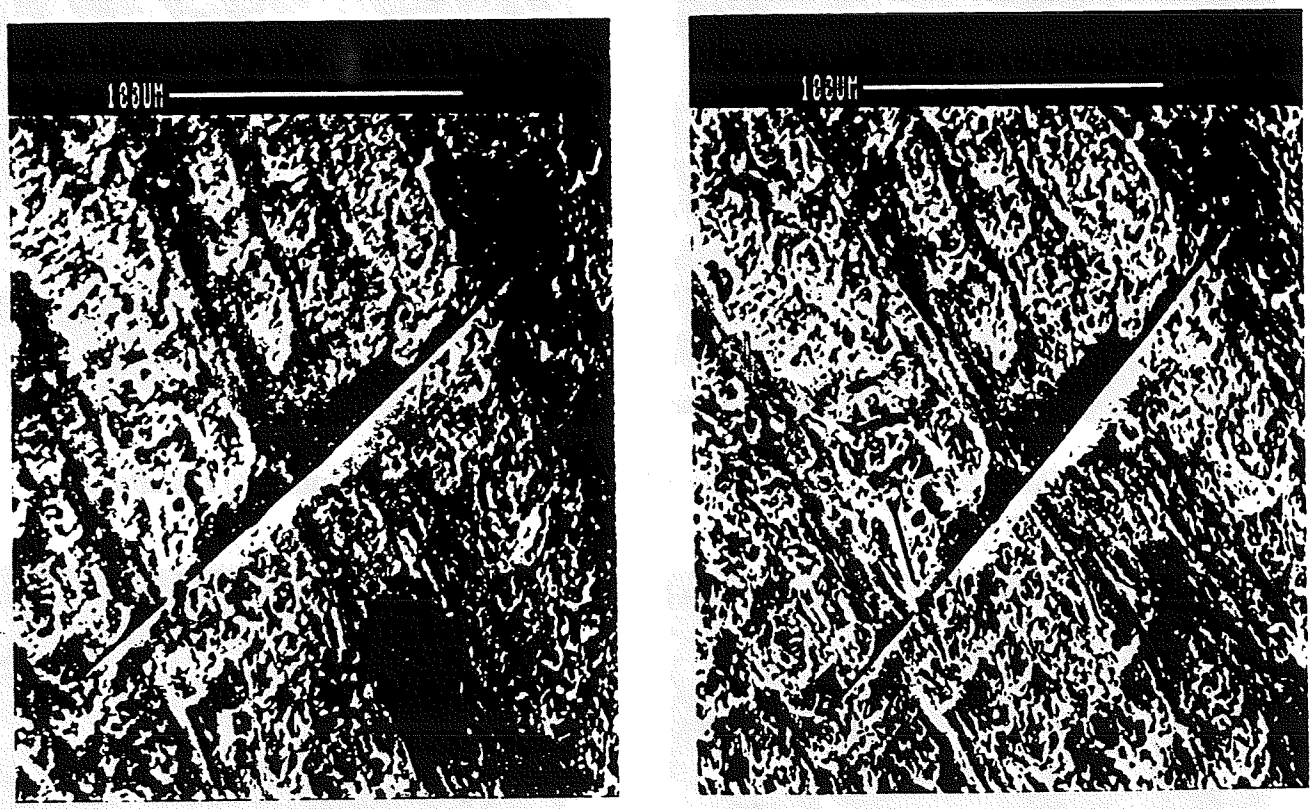


Figure 3-7. SEM of Titanium (sample Ti-29) in its initial condition (left) and after 20 ultrasonic cleaning cycles (right).

The aluminum coupons did exhibit an interesting phenomenon after 20 ultrasonic cycles. As shown in Figure 3-8, in the vicinity of surface irregularities, such as the engraved numbers, there were areas of cavitation damage to the surface, having the appearance of a "shadow". It is possible that the surface irregularities perturbed the local flow conditions and caused very localized cavitation. This phenomenon was seen only on the aluminum coupons.

The non-metal coupons were also viewed with an optical microscope and with the SEM but no surface changes were apparent.

3.1.3 X-Ray Photelectron Spectroscopy

One coupon of each material type was also examined using X-Ray Photoelectron Spectroscopy (XPS) initially and after 1 and 20 cleaning cycles. The XPS data is presented in Table 3-2. XPS examines only the top few atomic layers of the surface which includes a significant amount of ambient carbon. The large aluminum coupon (#AL5-01) and the large stainless steel coupon (#SS3-57) showed a reduction in the carbon levels after ultrasonic cleaning and an increase in the oxygen levels suggesting an increasingly oxidized surface or that the oxidized surface is being revealed by the removal of carbon. The other coupons were not analyzed after the twentieth cycle in a timely manor and may have picked up carbon again from the atmosphere.

3.2 Aluminum Foil Tests

3.2.1 Overall Evaluation

The aluminum foil sheets that were placed in the ultrasonic baths for a duration of three minutes were evaluated visually as to the extent of the damage caused by cavitation. Each sheet spanned the entire width of the bath (36" or 28") and the complete depth above the rack frame (18"). To summarize the damage, each sheet was divided into a grid of 4 x 4 inch squares and a number assigned to each square of 1 to 5 with 5 being the most severe damage. This data was then converted into gray-scale values and is presented in Figure 3-9. Each column of rectangles represents the family of sheets from one bath. Thus, the left-most column shows the six vertical slices from Bath A. The locations (A6, A5, etc.) were previously shown in Figure 2-1 and are approximately 10 inches apart.

As can be seen in Figure 3-9, the ultrasonic intensity of the baths was far from uniform. Bath A showed the highest intensity of the four baths which was not surprising since it was the only bath operating at 27 kHz. What is surprising is the variation in intensity within a given sheet, such as B1, or from sheet to sheet, such as B1 and B2. The foil sheet at B2 showed virtually no cavitation effects, yet it was only 10 inches from sheet B1 which showed the highest intensity in Bath B.

It is interesting to compare the intensity maps of Baths C and CR because both baths are operating at the same frequency (40 kHz) and the main difference between them is that Bath C contains Du Pont Zonyl™ FSN and Bath CR is a rinse bath containing only deionized water. Bath CR showed a much higher intensity than Bath C.

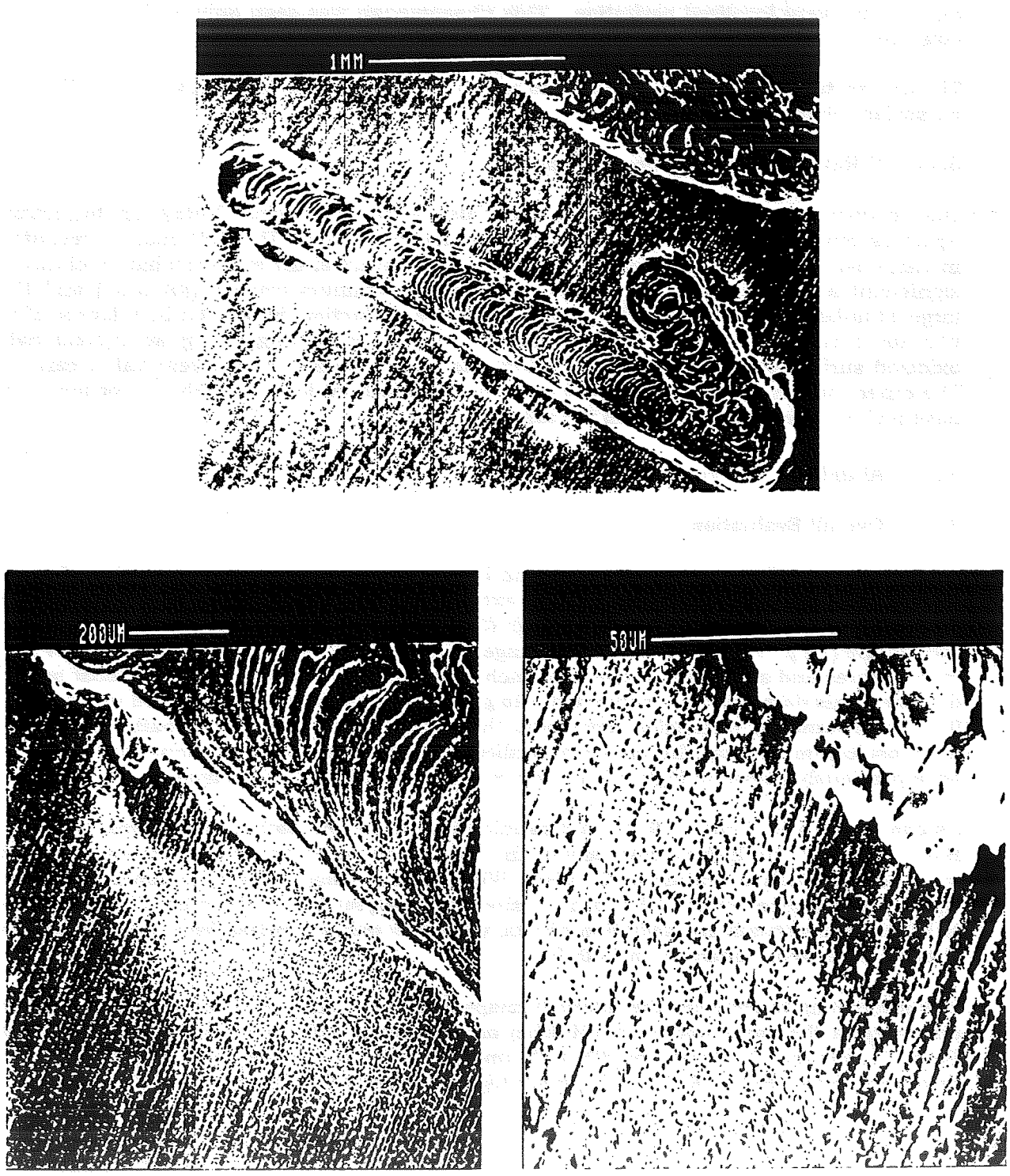


Figure 3-8. SEM of Aluminum (sample Al5-01) after 20 ultrasonic cleaning cycles showing cavitation damage in the vicinity of the engraved number "1".

Table 3-2. XPS Analysis of Coupons.

	Atomic concentration, %		
	Initial	1 cycle	20 cycles
Aluminum: AL5-01			
Carbon	30.75	22.39	21.21
Aluminum	18.97	21.40	20.37
Oxygen	50.28	54.28	58.41
Fluorine	0	1.94	0
Aluminum: A3			
Carbon	61.94	25.30	39.63
Oxygen	28.94	51.71	43.45
Aluminum	9.11	20.52	16.91
Fluorine	0	2.48	0.00
Stainless Steel: SS3-57			
Carbon	51.70	39.51	28.77
Oxygen	43.26	45.49	54.20
Chromium	5.04	7.29	10.00
Fluorine	0	4.34	6.37
Iron	0	3.37	0.66
Stainless Steel: A1			
Carbon	50.75	35.71	51.37
Oxygen	42.83	53.05	44.83
Chromium	3.43	5.70	3.80
Fluorine	0	4.00	0
Iron	2.99	1.54	0
Titanium: Ti-29			
Carbon	34.66	26.96	29.27
Titanium	10.77	15.19	21.15
Oxygen	53.50	57.32	46.11
Sodium	1.07	0	0
Lead	0	0.53	0.89
Fluorine	0	0	2.57
Vespel: A1			
Carbon	76.83	77.13	68.92
Nitrogen	4.69	4.68	4.41
Oxygen	15.09	13.77	19.59
Fluorine	3.40	4.43	7.08
Viton: V100			
Carbon	60.69	53.27	50.84
Oxygen	17.44	13.18	12.89
Fluorine	21.87	33.55	36.27

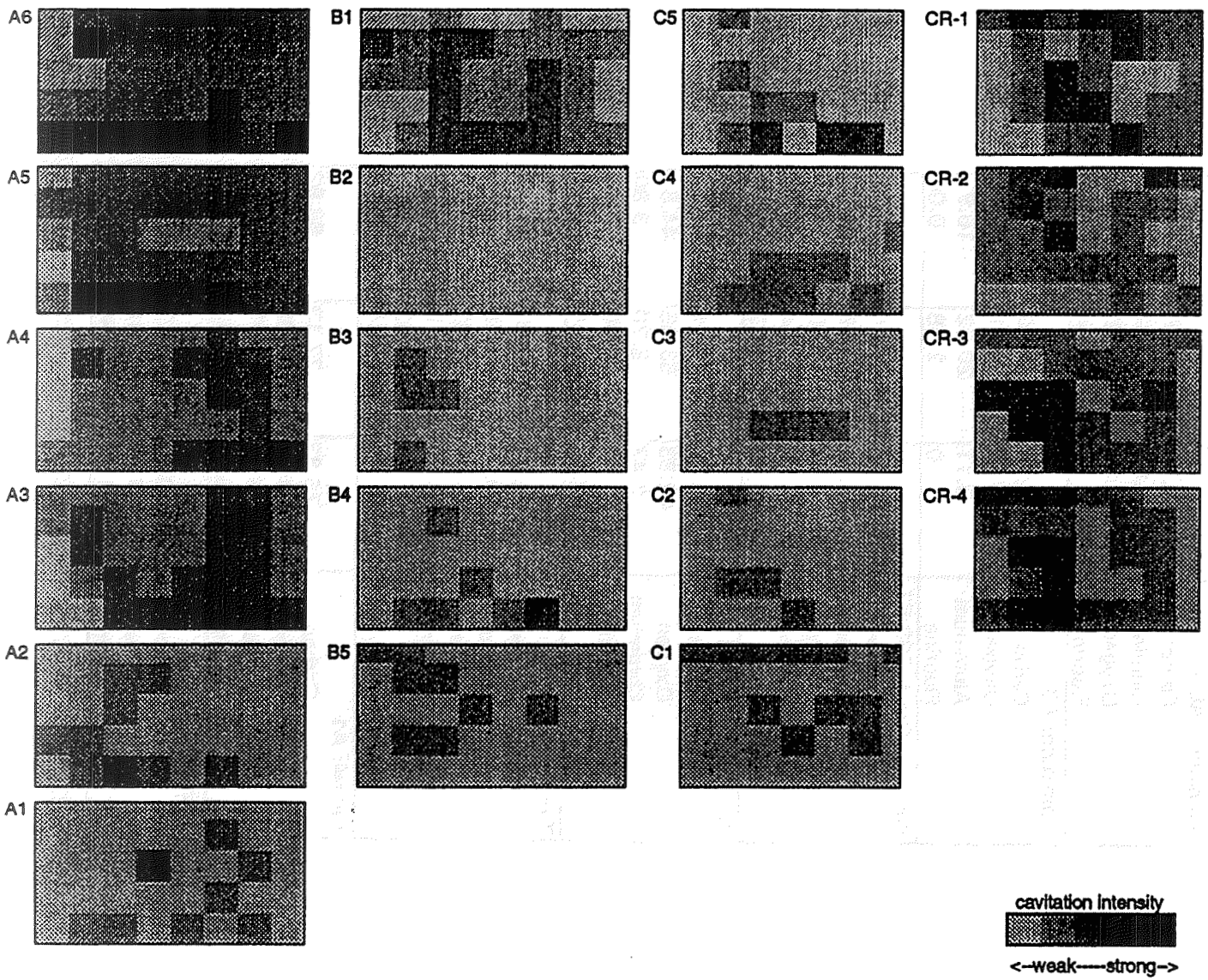


Figure 3-9. Ultrasonic intensity maps of baths as measured by aluminum foil cavitation damage. Each rectangle illustrates the damage to a sheet of aluminum foil placed vertically in the baths. For locations (A6, A5,...) see figure 2-1.

3.2.2 Localized Damage

The aluminum foil sheets, after exposure to the cavitation baths, were marked on the surface by dimples, dents, small and large holes, and crinkling. Figure 3-10 shows examples from sheet A5 of crinkling, which may have been caused by collapsed dents, and dimples which were often arranged in trails. In Figure 3-11 are seen examples of large dents on sheet A1 with small holes in the dents that were formed from both directions. Figure 3-11 also shows a magnified view of a crinkled area on sheet A3. Examples of small and large holes are shown in Figure 3-12. The single hole in the left frame of Figure 3-12 is approximately 320 microns in diameter and was caused by a single event such as a cavitation bubble bursting. The much larger hole in the right frame of Figure 3-12 (3 mm x 5 mm) was formed by the agglomeration of many smaller holes in close proximity with the small holes going in both directions.

Figure 3-13 shows an interesting phenomenon that was observed several times on sheet C5: a circular pattern of dimples, some of which have burst, and all are going in the same direction with the foil in the center of the circle forming a dome in the opposite direction of the dents. The sample was tilted at a 45° angle for the photograph. Additional examples of holes in the aluminum foil from sheet CR2 are seen in Figure 3-14. The small holes were about 100 microns to 200 microns in diameter with a larger hole again formed by the agglomeration of smaller ones.

3.2.3 Standing Wave Patterns

Evidence of the presence of standing wave patterns were observed on some of the aluminum foil sheets from two of the ultrasonic baths: B and CR. The other two baths, A and C, contained surfactants, Brulin™ and Zonyl™ respectively, which may dampen or disturb the wave patterns in some way. Figure 3-15 is a photograph of one entire sheet of aluminum foil (#CR3), measuring 18" x 28", illuminated from behind to show the pattern of holes. As can be seen in the figure, the holes are clustered together and the clusters are spaced in relatively regular intervals. The ultrasonic waves in the water should obey the following equation:

$$C = \lambda \nu$$

where C is the velocity of sound in water, λ is the wavelength, and ν is the frequency. The velocity of sound in water is 1594 m/s at 65°C (149°F) [Reference 3] and the frequency in Bath B and CR is 40 kHz. Substituting these values into the above equation, one can calculate a theoretical wavelength of 40 mm. Measurements were taken of the spacing between several of the periodic clusters of holes on sheets from Baths B and CR, and assuming the spacing between two adjacent clusters is one-half a wavelength, experimental values of the wavelengths were obtained and are summarized in Table 3-3. As seen in Table 3-3, the agreement between theory and experiment is excellent.

Table 3-3. Ultrasonic Wavelength Comparisons

	λ = wavelength
Theoretical	40 mm
Experimental - Bath B	38 mm
Experimental - Bath CR	41 mm

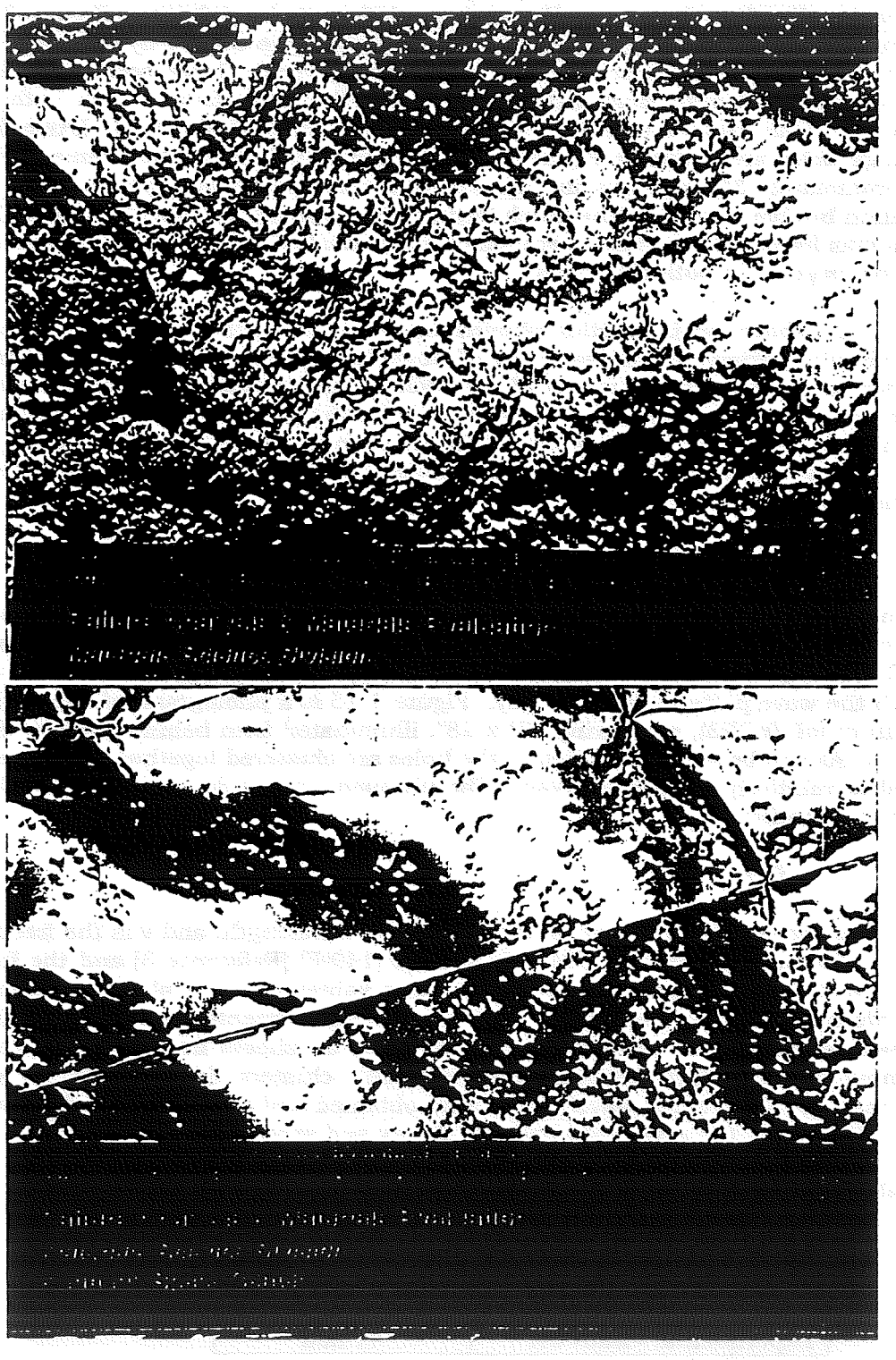


Figure 3-10. Examples of damage to aluminum foil #A5: crinkling (above), and trails of dimples (below).

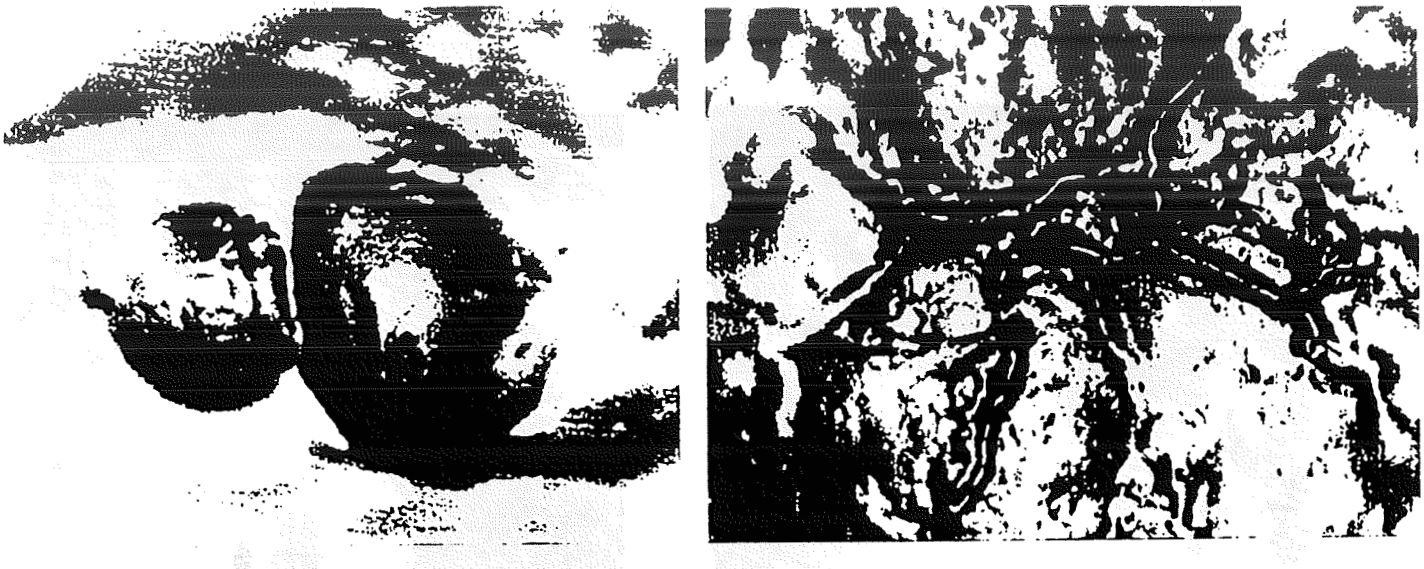


Figure 3-11. Examples of damage to aluminum foil: large dents with small holes in foil #A1 (left); crinkling on foil #A3 (right). Magnification is 6.3x.

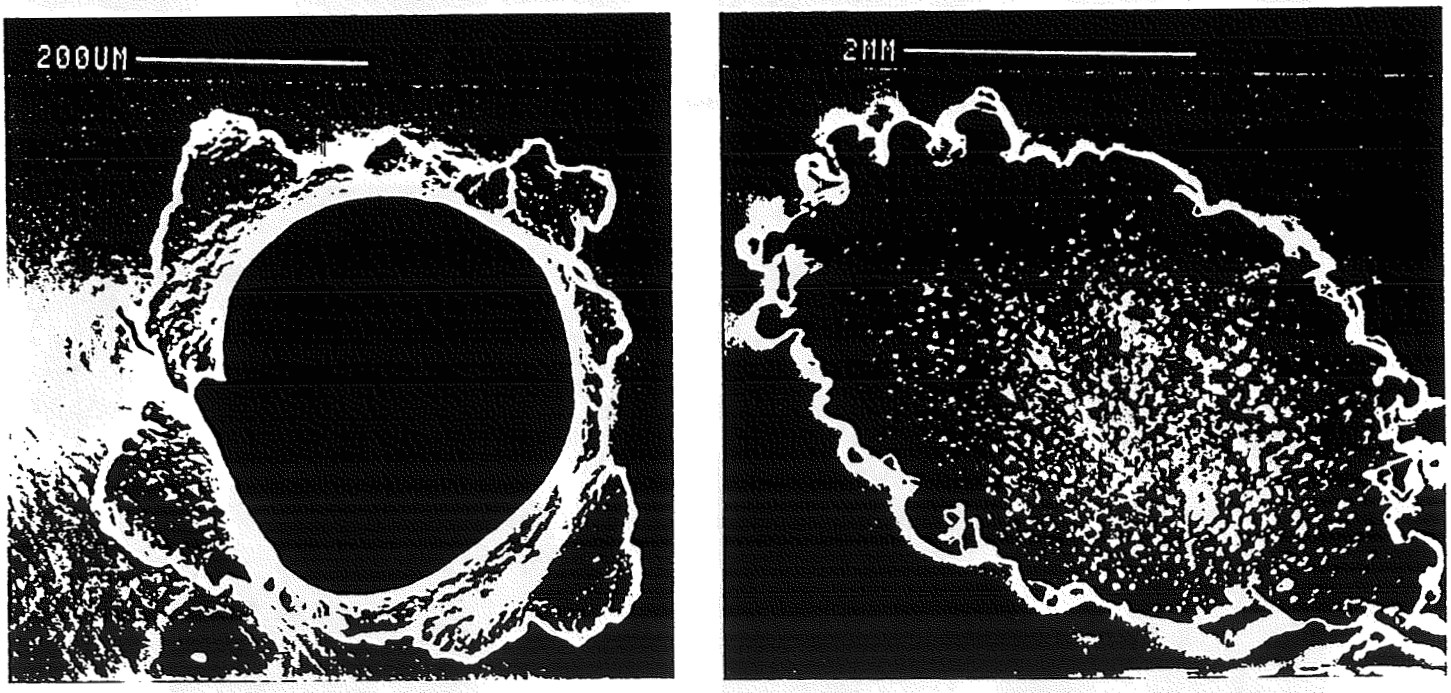


Figure 3-12. Examples of holes in aluminum foil #A3: single hole, diameter 320 microns (left); large hole, diameter 5 mm, resulting from numerous smaller holes (right).

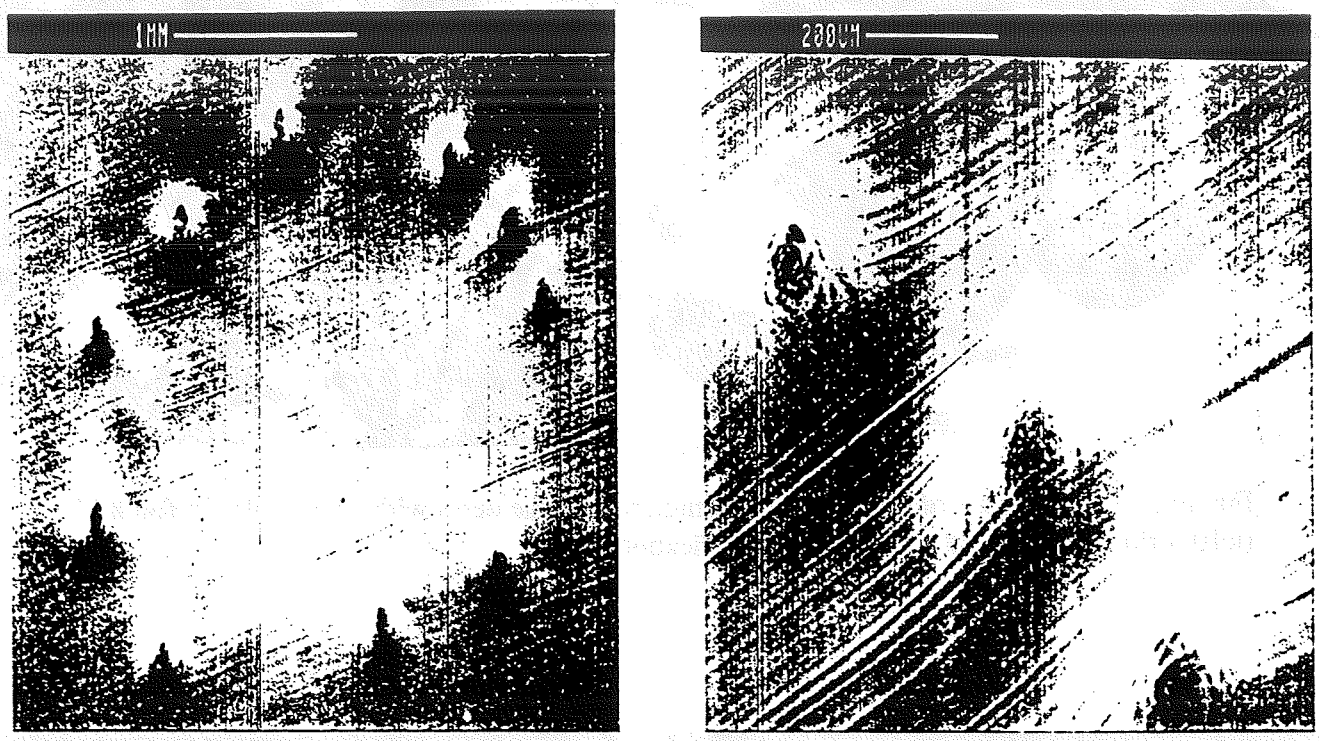


Figure 3-13. Examples of holes in aluminum foil #C5: the holes and dimples are arranged in a circle (left); the holes and dimples are oriented in the same direction (right).

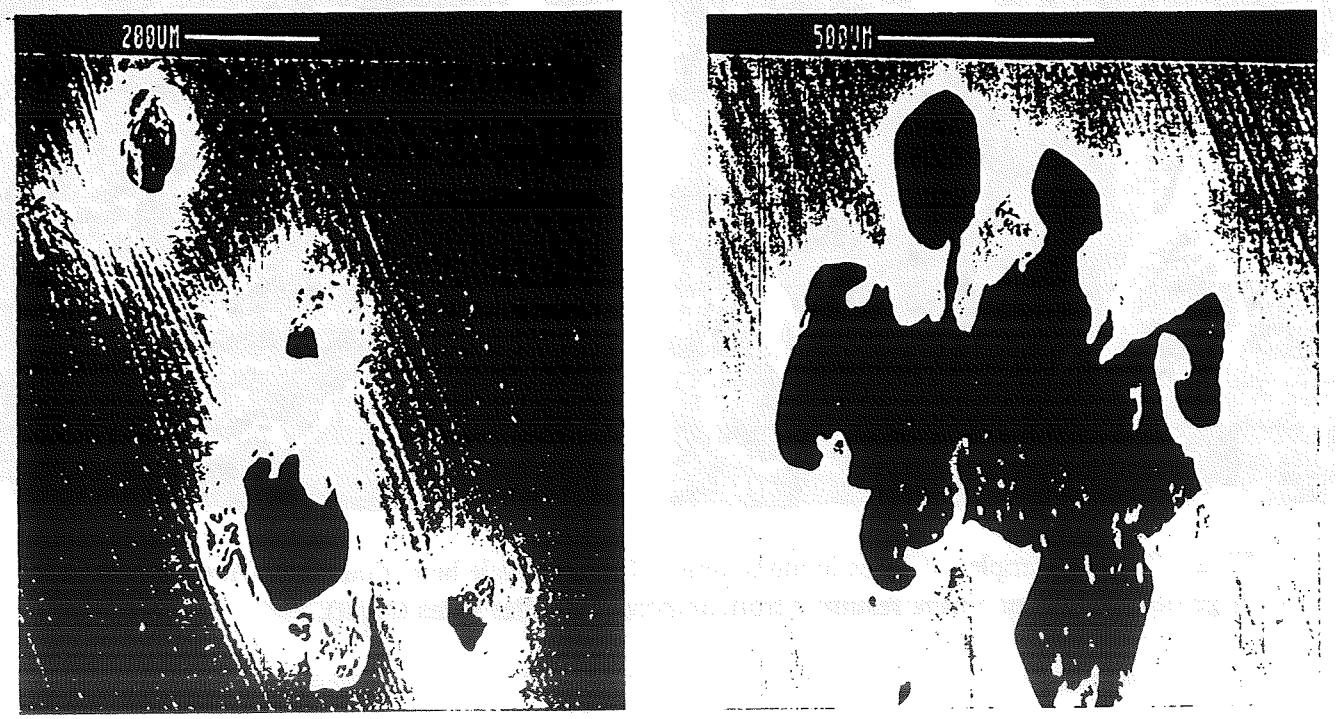


Figure 3-14. Examples of holes in aluminum foil #CR2: holes going in both directions (left); large hole formed by several smaller holes (right).

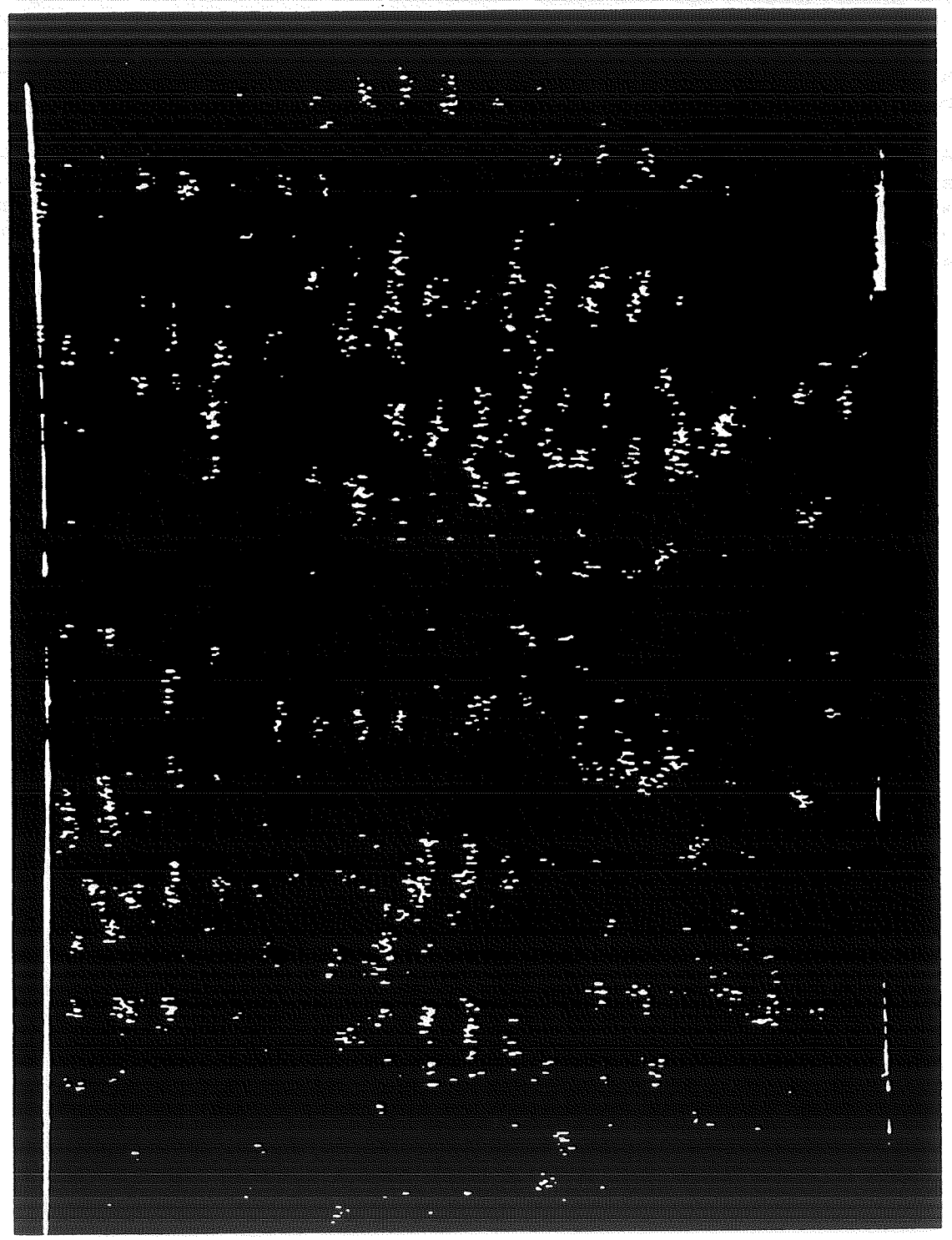


Figure 3-15. Backlight photograph of aluminum foil #CR3 showing hole patterns.

IV. CONCLUSIONS

The primary focus of this work was to evaluate the effect of the ultrasonic cleaning process at CRCA on the surfaces of metal and non-metal coupons. The mass loss rate from the metal coupons after 20 cycles of ultrasonic cleaning was very low: 0.11% or less for aluminum, 0.0019% for titanium, and 0.0014% for stainless steel. The non-metal coupons, Vespel™ and Viton™, showed some periods of weight gain as well as weight loss.

A secondary focus of this work was to characterize the intensity of the ultrasonic baths as a function of location. The primary motivation for this was to identify possible areas of low activity to be avoided in the coupon study. It was found that the ultrasonic baths were not uniform. The two baths that did not contain a surfactant showed evidence of standing waves that agreed with the theoretical value for the wavelength.

V. REFERENCES

1. Ensminger, Dale, Ultrasonics, Fundamentals, Technology, Applications, Marcel Dekker, Inc., New York, 1988.

2. Mehta, N.H., "Evaluation of Ultrasonic Cavitation of Metallic and Nonmetallic Surfaces", NASA/ASEE Summer Faculty Fellowship Program, NASA CR-191004, September, 1992, pp. 258-315.

3. Weast, Robert C., editor, CRC Handbook of Chemistry and Physics, 46th ed., The Chemical Rubber Co., Cleveland, Ohio, 1965, p.E-29.

1995 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAMM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

511-37
7751
P. 33
(34)

DESIGN OPTIMIZATION OF A BRUSH TURBINE WITH A CLEANER/WATER BASED SOLUTION

Dr. Rhyn H. Kim
Professor
Department of Mechanical Engineering and Engineering Science
University of North Carolina at Charlotte
Charlotte, North Carolina

KSC Colleague - Rudy Werlink
Fluids

Contract Number NASA-NGT-60002
Supplement 19

July 21, 1995

PAGE 304 INTENTIONALLY BLANK

ACKNOWLEDGEMENT

I would like to acknowledge that I have been privileged to participate in the FACULTY SUMMER PROGRAM in KENNEDY SPACE CENTER sponsored by NASA in 1994 and 1995. I appreciate the assistance offered by the Coordinator staff, Drs.Anderson and Hosler particularly help in the area of the paper work given by Ms. Kari Stiles of University of Central Florida.

I have enjoyed working with Mr.Rudy Werlink, KSC colleague who has helped me whenever I needed in the real engineering and office oriented problems. I thank Dr. Gary Lin, former chief of the branch for his consistent interest on the subject I have spent time for two years.

The last but not least, I acknowledge my office mate, Prof. Pao-lin Wang for his continuous encouragement, and a moral support given to me during the tenure of the assignment in KSC.

ABSTRACT

Recently, a turbine-brush was analyzed based on the energy conservation and the force momentum equation with an empirical relationship of the drag coefficient. An equation was derived to predict the rotational speed of the turbine-brush in terms of the blade angle,number of blades,rest of geometries of the turbine-brush and the incoming velocity. Using the observed flow conditions , drag coefficients were determined. Based on the experimental values as boundary conditions,the turbine-brush flows were numerically simulated to understand first the nature of the flows, and to extend the observed drag coefficient to a flow without holding the turbine-brush.

SUMMARY

One of methods of cleaning fouling in tubes will be a turbine-brush unit to replace R113 as a cleaner after 1995. The turbine-brush consists of a turbine and a brush with a connector between the two. The turbine-brush was created in KSC,NASA and demonstrated its rotational motion while it was held at a position.

Objective of the program is to optimize the geometry of the turbine-brush in terms of the blade angle, the number of blades, and arrangements of the connector and brush, flow rates, rotational speeds, diameters of tubing and performance of cleaning before a final product is designed and manufactured.

A prediction of a rotational speed was made without considering the blade angle in the beginning stage of the project and later the blade angle was taken into account. In the analysis of rotation speeds, empirical relations were used for drag forces. The drag forces were determined in a situation where the turbine-brush was held in a position. One suspects whether the drag force measured in this way would be the same as the unit moves. Because of this reason, a numerical simulation was thought of as a method of understanding the nature of the flow in a turbine-brush flow and of determining a way to use the simulation technique to a freely moving turbine-brush unit i tubes.

NEKTON was available for analyzing a simulation model. A simplified model, and an improved model for the blade section were made and these models were being tested. It seems that a superposition method of two solutions, one out of a rotating model and the second one of that model without rotation in a turbulent flow should be sought for a final simulation of the turbine-brush flows.

TABLE OF CONTENTS

LIST OF FIGURES

NOMENCLATURE

A	Area of a flow cross-section, m^2
a	Constant used in the boundary conditions
b	Constant used in the boundary conditions
$C_{1\varepsilon}$	Constant used in the boundary conditions
$C_{2\varepsilon}$	Constant used in the turbulence model
$C_{3\varepsilon}$	Constant used in the turbulence model
C_{μ}	Constant used in the turbulence model
c	Constant used in the boundary conditions
d	Diameter of the tube, m
E	Available energy for rotation Joule/m^3
F(x,t)	Function used in the boundary condition
Fi	Body force in the i-the direction
G_b	Function used in the turbulence due to buoyancy
G_{κ}	Function used in the turbulence due to dissipation kinetic energy
gi	Gravitational acceleration, m^2/s
J_{total}	Sum of the moment of inertia of rotating elements
mdot	Mass rate, Kg/s
p	Static pressure, pa
Q	Volume rate along the blades, m^3/s
S_m	Mass source of dispersed states
t	time,s
$(\Delta t)^2$	Convergence criterion
u_1	Tangential velocity component in the y-direction
u_i	Velocity component in the i-the direction
u_n	Velocity component to normal to the surface
u_x	Velocity component in the x-direction
u_y	Velocity component in the y-direction
u_z	Velocity component in the z-direction
u_iu_j	Reynolds stress
V	Velocity, m/s
W_{drag}	Rate of work of drag
$W_{kinetic}$	Rate of work due to rotation
x_i	coordinates

Greek Letters

β	Angle between the relative velocity and tangential velocity
δ_{ij}	Kronecker delta
ε	Dissipation energy in turbulence
κ	Dissipation kinetic energy
μ	Molecular viscosity
μ_t	Turbulent viscosity
ν_{eddy}	Turbulent kinematic viscosity
ρ	Density of the fluid
τ_{ij}	Shear stress tensor
σ_h	Constant defined in the turbulence model
$\sigma\varepsilon$	Constant defined in the turbulence model
ω	Angular velocity

I. INTRODUCTION

CFC refrigerants/cleansers will not be allowed to use in cleaning fouling in tubes starting in 1996. A mechanical device was devised to clean inner surfaces of tubes in order to cope with the situation in which R113 is not available as a cleanser. An analytical expression of an angular velocity of the turbine-brush was developed in terms of geometries of the unit, the blade angle, number of blades, and the incoming velocity of the flow. The expression includes a drag coefficient which was derived from a series of experiments from an open loop and a closed loop experimental set-ups, respectively.

A numerical simulation was conducted first to understand the nature of the turbine-brush flows as the unit rotated while it was held at a position, and second to determine whether the numerical simulation may be extended to a case in which the turbine-brush moves forward while it rotates. For this purpose, a simplified model was developed and tested. An improved numerical model for the blade was also developed for further test.

At present, a numerical simulation with the simple model has not been completed. The test data from the closed loop set-up have not been analyzed, yet. The parts which are not completed will be further looked into on campus of UNC-Charlotte.

II. APPROACH TO A SIMULATION

A numerical simulation was conducted for the turbine-brush flows by NEKTON of FLUENT Inc.

NEKTON is based on a finite element method incorporated with a spectral element technique originated by Patera(1984). This scheme employs a five point interpolating polynomial, known as the Legendre polynomial on a finite element network. As a result of the higher order power series, the software produces a fast converging numerical result with a relatively crude element network. However, this works for Stokes flows of a low Reynolds number up to 12,000 with a K-ε turbulence model(Fluent,Inc.,1994).

The results of the numerical simulation will be compared with the results of experiments in terms of pressure and velocity distributions. The results of the lower Reynolds number flows may not represent the real flow conditions at all because of simplifying assumptions to build a model for simulation. In the beginning of simulation procedures, a relatively easy construction of the grid network and low rotational speed of the turbine brush lead to a use of NEKTON. Simplifying assumptions will be explained in later sections.

The procedure and result of the simulation analysis give a first hand insight in generating the elements and flow patterns. Application of this sort, a brush in the region of wake generated by a turbine, did not exist in a real world. The turbine brush was invented and its patent application was submitted in November 1993.

2.1 Geometric Model

A simplified geometric model is explained as follows. The simplified model is a first step for an eventual success of the simulation. Rather than the transient rotation of the turbine, steady angular velocity or a tangential velocity is assumed for the turbine-brush. Four blades are placed normal to the direction of the flow 90° apart in the outer race of a bearing. The bearing is located at the center of a blade housing. A wire is attached to the center of the bearing in such a way that the turbine-brush can rotate being held in a position. The liquid turbine and the rotating part of the bearing are connected to the brush by a two bar linkage, consequently make the brush rotating. The brush cleans fouling from the surface of the tube.

The first section of the simplified geometric model is the entering section of a fluid. The distance of this section is 1d, d being the diameter of the tube, since an incoming velocity of the flow is uniform. The second section consists of the entire structure of the turbine including the ring, blades, and bearing. The ring, blades and bearing are simplified to be a circular plate with four radial bars on each side. The turbine ring holds an O-ring so that the fluid may not pass the blade section in the direction of tube axis but moves in radial direction through the ring section of the turbine. The turbine runs without the O-ring. The third section is the connecting linkage. The fourth section is the brush section including the brush hub and a cut-off brush in four places separated equal distance along the circumference. It is assumed that fluid does not pass at all through the brush. The last two sections are the flow leaving sections which will be of a distance of (5 to 8)d. A sketch of the turbine-brush is shown in Figure 1. A discretization of the flow domain is illustrated in Figure 2.

2.2 Construction of a NEKTON Network

A main idea of finite element construction in a three dimensional flow domain for NEKTON is to make a two-dimensional element on a plane and extend it along the third axis. Finite elements should have four vertices on the plane, and the four vertices should be specified in the counter-clockwise direction. Since a five point interpolating scheme is used between two connecting points, a crude initial network is affordable. The element network can use global refinement techniques, such as 'zipper','picture frame','corner' and so forth. Thereby, an easier construction of elements is possible. The number of elements of an

initial model is a little over one hundred and are shown in Figures 3-8. Figure 9 shows a sketch of a cross-section of the blade cut on the mid section of the flow. It also shows the steps of constructing the improved model for the blade by taking a thinner segment along the direction of the flow with the help of the global refinement of NEKTON procedures. Pressure and velocity boundary conditions on different elements are illustrated in Figures 10 and 11.

Starting with Figures 12a through 12f, the continuously increasing blade sections are shown for the improved model.

III.GOVERNING EQUATIONS

A system of equations consists of continuity, and momentum equations. Since the temperature variations are not considered in this investigation, the energy equation is not needed.

3.1 Continuity

Mass conservation law, known as the continuity equation is:

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x_i}(\rho u_i) = S_m$$

where ρ is the density, t is the time, u_2 are the velocity components, and x_is are the coordinate variables.
Equation (1) is the general form of the mass conservation equation and valid for incompressible as well as compressible flows. The source S_m is the mass added to the continuous phase from the dispersed second phase,i.e., due to vaporization of liquid droplets.

3.2 Equation of Motion

Conservation of momentum in the i-th direction in an inertial reference frame is described by(Schlichting,1960):

$$\frac{\partial}{\partial t}(\rho u_i) + \frac{\partial}{\partial x_j}(\rho u_i u_j) = -\frac{\partial p}{\partial x_i} + \frac{\partial \tau_{ij}}{\partial x_j} + \rho g_i + F_i$$

where p is the static pressure, τ_{ij} is the stress tensor(described below), and g_i and F_i are gravitational acceleration and external

body forces in the i direction, respectively.

The stress τ_{ij} is given by:

$$9\tau_{ij}=[\mu(\frac{\partial u_i}{\partial x_j}+\frac{\partial u_j}{\partial x_i})]-\frac{2}{3}\mu\frac{\partial u_k}{\partial x_k}\delta_{ij}$$

where μ is the molecular viscosity and the second term on the right hand side is summed over all three component directions. In the computation, the term $(2/3)\mu(\partial u_k/\partial x_k)$ is ignored under the assumption that the divergence of velocity has a negligible effect on the stresses.

3.3 Model for Turbulent Flows

Modeling of turbulent flows requires appropriate modeling procedures to describe the effects of turbulent fluctuations of velocity and scaler quantities on the basic conservation equations presented in the previous section. The τ-ε turbulence model is a basic model from which a turbulence is described. This model is an eddy-viscosity model in which the Reynolds stresses are assumed to be proportional to the mean velocity gradients, with the constant of proportionality being the turbulent eddy viscosity, μ_t. This assumption, known as the Boussinesq hypothesis, provides the following expression for the Reynolds stresses(Fluent,Inc.,1994):

$$\rho\overline{u_iu_j}=\rho\frac{2}{3}\kappa\delta_{ij}-\mu_t(\frac{\partial u_i}{\partial x_j}+\frac{\partial u_j}{\partial x_i})+\frac{2}{3}\mu_t\frac{\partial u_l}{\partial x_l}\delta_{ij}$$

Here κ is the turbulent kinetic energy:

$$\kappa=\frac{1}{2}\sum_i\overline{u_i^2}$$

The term $(2/3)\mu_t(\partial u_l/\partial x_l)$ is ignored under the assumption that the divergence of velocity has a negligible effect on the turbulent stress tensor. The equation for the Reynolds stresses is analogous to that describing the shear stresses that arise in laminar flow

with the turbulent viscosity μ_t playing the same role as the molecular viscosity μ. Therefore the form of the turbulent momentum equations remain identical to the form of the laminar momentum equation except that μ is replaced by an effective viscosity, $\mu_{eff} = \mu + \mu t$. The turbulent viscosity μ_t is obtained by assuming that it is proportional to the product of a turbulent velocity scale and length scale.

In the κ-ε model, these velocity and length scales are obtained from two parameters: κ, the turbulent kinetic energy and ε, the dissipation rate of κ. The velocity scale is taken to be κ=5 and the length scale is taken to be $\kappa^{3/2}/\varepsilon$. Hence, μ_t is given by:

$$\mu_t = \rho C_\mu \frac{\kappa^2}{\epsilon}$$

where C_μ is an empirically derived constant of proportionality (set to a default value of 0.09). The values of κ and ε are obtained by solution of conservation equations (Fluent, Inc., 1994):

$$\frac{\partial}{\partial t}(\rho\kappa) + \frac{\partial}{\partial x_i}(\rho u_i \kappa) = \frac{\partial}{\partial x_i}\left(\frac{\mu_t}{\sigma_k}\frac{\partial \kappa}{\partial x_i}\right) + G_\kappa + G_b - \rho\epsilon$$

$$\frac{\partial}{\partial t}(\rho\epsilon) + \frac{\partial}{\partial x_i}(\rho u_i \epsilon) = \frac{\partial}{\partial x_i}\left(\frac{\mu_t}{\sigma_\epsilon}\frac{\partial \epsilon}{\partial x_i}\right) +$$

$$C_{1\epsilon}\frac{\epsilon}{\kappa}(G_\kappa + (1 - C_{3\epsilon})G_b) - C_{2\epsilon}\rho\frac{\epsilon^2}{\kappa}$$

where $C_{1\varepsilon}$ and $C_{2\varepsilon}$ are empirical constants, σ_κ and σ_ε are "Prandtl" numbers governing the turbulent diffusion of κ and ε, G_κ is the rate of production of turbulent kinetic energy:

$$G_k=\mu_t\left(\frac{\partial u_j}{\partial x_i}+\frac{\partial u_i}{\partial x_j}\right)\frac{\partial u_i}{\partial x_j}$$

and G_b is generation of turbulence due to buoyancy:

$$G_b=-g_i\frac{\mu_t}{\rho\sigma_h}\frac{\partial \rho}{\partial x_i}$$

The above equations are applied to minimizing energy principle, in return, one obtains a series of finite element based equations. These equations are solved under physical boundary conditions. Actually, these boundary conditions are derived from experiments.

IV.BOUNDARY CONDITIONS

Two types of boundary conditions for elements exist in the flow domain in consideration. The first is on the boundary surface of each element, and the second is on the element itself. There are also conditions of the inlet and exit of the flow. Variables involved with the boundary conditions are the pressure and velocity. The first condition we consider is an incoming velocity. Unless otherwise specified, the incoming velocity is fully developed, meaning that either a uniform speed or a profile of the velocity of the flow is known. A simulation analysis yields the velocity at the exit of the flow field. Variations of velocity at the exit are usually shown in the region of 7 to 9d downstream of the inlet. These are identified VELOCITY and OUTFLOW.

4.1 Pressure Boundary

The pressure is reduced as the fluid flows due to friction and changes in cross-sections of the flow passages. In general, this is expressed as:

$$\partial p/\partial x + F(x,t)= 0$$

where p is the static pressure, x, the coordinate in the flow

direction, t, the time and F is a function in x and t.

4.2 Boundary Conditions for Velocity

The fluid is in contact with a stationary surface(inner surface of conduits) and with a moving surface(blades,ring structure, connector, and brush). In continuum condition, the velocity does not exist at the stationary surface:

$$u_x = 0,\ u_y = 0,\ \text{and}\ u_z = 0$$

On the other hand, at the moving boundary, the fluid layer next to the surface, moves with the same speed as the surface and expressions of velocity components are given by:

$$u_n = 0,\ u_1 = 0,\ u_2 = r\omega{=}\omega(y^2 + z^2)^{0.5}$$

where u_n is the component of the velocity normal the surface, u_1 is the first tangential(y-axis) component, and u_2 is the second tangential component(z-axis) of the velocity.

For each element of the fluid in the field, the above boundary conditions entered. A partial list of elements and their boundary conditions are shown in Appendix starting with the element 1 through the element 4 and element 328. In the list, initial conditions of the velocity is also provided.

Specifically, the current project uses the following for the above expression.

Incoming velocity, $u_x = f(x)$ or constant, $u_y = 0$, and $u_z = 0$

Pressure, $pa=a - bx$

Moving Layer, $u_n = 0,\ u_1 = 0,\ u_2 = c*(y^2 + z^2)^{0.5}$

where a, b, and c are determined based on the specific runs of experiments. With an existing set of experimental data, the list shows those specific numbers. A sinusoidal equation is assigned for u_x to indicate that the at the end of the turbine-brush unit, u_x vanishes, implying that only the rotational speed influences the flow domain.

4.3 Body Force

The force holding the turbine-brush at a point is considered a body force opposite to the flow direction.

V.MATERIAL PROPERTIES

The fluid in consideration is liquid water at room conditions. Therefore, variables in mind are the density and viscosity which are constant because the temperature variation is not considered.

VI.SOLUTION METHOD

NEKTON provides a unsteady solution for three dimensional Navier-Stokes problems. The convective terms are treated explicitly by using a third order Adams-Bashforth multiple step scheme which is accurate to $O(\Delta t^2)$. The diffusion terms are treated implicitly by using a third order backward differentiation multistep scheme.For problems with time-dependent domains, the geometry is updated explicitly using a third order Adams-Bashforth multistep scheme.

Once an element network is constructed, an input file is submitted into the SOLVER of NEKTON. Based on error messages, the number of elements or of orders may be increased.

6.1 Current status of Computations

The first solution was tried with the number of elements of about 150. The NEKTON solver did not provide a converging solution. By using the global refinement techniques, the number of elements were increased to the neighborhood of 450. The upper limit of the three dimensional elements is 450. Thus, a consistent attention must be paid in a process of building elements.

There are two versions of the blade models. At present, the simplified version has been investigated. Because of a few reasons, a solution has not been produced yet. In order to determine whether the geometry of the model is possible at all, a flow of Reynolds number of 100 was tried. A result was produced. Therefore, the model used for simulation is possible.

Further trials will be made on campus of UNC-Charlotte. The trials will include the first and second models of the blade section as well as establishing the steady flows for the both models.

VII. MODIFIED EQUATION FOR PREDICTING ROTATIONAL SPEEDS

Shepherd(1967) gives the required rotational energy in terms of an available energy for rotation as follows:

$$E = u_2(u_2 - Q/A_2 \cot\beta_2) - u_1(u_1 - Q/A_1 \cot\beta 1)$$

where Q is the volume rate flowing along the blades, A, the cross-section of the flow and the angle between the relative velocity of the flow rate and the tangential speed u is denoted as β(see Figure 13). Subscripts 1, and 2 refer to an entering and leaving sections, respectively. Applying the above equation to the turbine-brush with $\beta_1 = \pi/2$, one obtains the required energy for rotation as:

$$E = u_2(u_2 - Q/A_2 \cot\beta_2)$$

7.1. An Equation for the Angular Velocity

The existing turbine-brush has $\beta_2 = \pi/4$. Based on report(Kim,1994 and Kim and Werlink,1995), the conservation energy equation is given by:

$$W_{drag} + W_{kinetic} = mdot*[(p_{exit} - p_{inlet}) + (V_{exit}^2 - V_{inlet}^2)/2$$

where W_{drag} is the rate of work due to the drag force and friction, $W_{kinetic}$ is the minimum kinetic energy for the turbine-brush to rotate and is expressed as:

$$W_{kinetic} = J_{total} \cdot \omega^2/2$$

where J_{total} is the sum of moment of inertia of the entire rotating body and mdot is given by:

$$mdot = \rho\ V_{inlet}\ A_{tube.cross}$$

Therefore, a relation for the angular velocity ω is obtained by optimizing the energy equation with respect to ω under the constraints, the mass rate ≥ 0 and

$$E \geq W_{drag} + W_{kinetic}$$

VIII.EXPERIMENTS WITH A CLOSED LOOP APPARATUS

A closed loop apparatus for the turbine-brush and its instrumentation were completed very late. Specially, a ultra-sonic volume rate reading device did not function properly, and a series of second testings were done again. Instrumentations were described elsewhere(Kim,1994 and Kim and Werlink,1995). The loop with the test section is shown in Figure 14. A quick inspection of the current data indicated that there are variations between the data of an open loop and closed loop systems. A careful investigation is necessary for the new set of data before any comparison is made between the current and old data sets.

IX.CONCLUDING REMARKS

A numerical simulation of the turbine-brush flows was conducted with a simplified model and computations are on going. After a successful result comes from the simplified model, the improved blade model will be simulated. A series of new data sets were obtained from a closed loop fluid circuit. An analysis of the new data will be done.

Unfinished portions of the project will be reported as soon as possible.

X.REFERENCES

Fluent,Inc.,1993 NEKTON User' Guide

Kim,R.H.,1994,Design optimization of a brush turbine with a cleaner/water based solution, 1994 NASA/ASEE summer faculty fellowship program, University of Central Florida,NASA-NGT-60002Supplement:17,pp.313-338

Kim,R.H. and Werlink,R.J.,1995,Analysis of a turbine-brush for cleaning fouling in tubes, Proceedings of Fluid Engineering, ASME,Fluid Machinery Forum,FED222,pp.93-99

Patera,A.T.,1984,Spectral Element Method for Fluid Dynamics:Laminar Flow in a Channel Expansion, J. of Computational Physics,Vol.53(3),pp.468-488

Schlichting,H.,1960,Boundary Layer Theory,McGraw-Hill Book Co.

Shepherd,D.G.,1967, Principles of turbomachinery, The Macmillan Co.,pp.65-74

X.APPENDIX

Abridged Input File

```
   2.90000 NEKTON VERSION
 3 DIMENSIONAL RUN
 55 PARAMETERS FOLLOW                                          19
 1000.00          DENSITY
0.108000E-02      VISCOS
      0.          BETAG
      0.          GTHETA
      0.          P5
      0.          P6
 1.00000          RHOCP
 1.00000          CONDUCT
      0.          QVOL
      0.          FINTIME
 1.00000          NSTEPS
      0.          DT
      0.          IOCOMM
      0.          IOTIME
      0.          IOSTEP
      0.          MAXHIS
      0.          PATCHMESH
0.500000E-01      GRID
-1.00000          INTYPE
 5.00000          NORDER
      0.          DIVERGENCE
0.100000E-03      HELMHOLTZ
      0.          NPSCAL
0.100000E-01      TOLREL
 1.00000          TOLABS
0.250000          COURANT
 2.00000          TORDER
      0.          TMESH
      0.          MAXCGHELM
      0.          PoPRESSURE
      0.          GAS CONS R
      0.          Lo Ma GRAV
 100.000          MXGEOM
 100.000          MAXNEW
0.100000E-03      TOLDIR
0.100000E-03      TOLNEW
0.100000E-01      TOLVNORMAL
      0.          SSFSLIMIT
      0.          P39
      0.          IFMAGNET
      0.          SIGMAMAG
      0.          SSFSLENG
 1.00000          VHScaleFac
      0.          POISSONRA
      0.          FSDTFAC
      0.          IRSTART
      0.          CNFAC
      0.          PGEOM
0.140000          MIXLFAC
      0.          MAXCGFS
      0.          SETSIGMA
 4.00000          REMESH
      0.          PRFSDATA
      0.          RAMPSTEPS
      0.          VRELAXFAC
    4  Lines of passive scalar data follows2 CONDUCT; 2RHOCP
 1.00000       1.00000       1.00000       1.00000       1.00000
 1.00000       1.00000       1.00000       1.00000
 1.00000       1.00000       1.00000       1.00000       1.00000
 1.00000       1.00000       1.00000       1.00000
 18  LOGICAL SWITCHES FOLLOW
 T     IFFLOW
```

```
  F      IFHEAT
  T      IFTRAN                                                          20
  T  F  F  F  F  F  F  F  F  F  F IFNAV & IFADVC (convection in P.S. fields)
  F  F  T  T  T  T  T  T  T  T  T IFTMSH (IF mesh for this field is T mesh)
  F      IFAXIS
  T      IFSTRS
  F      IFSPLIT
  F      IFMGRID
  F      IFMODEL
  F      IFKEPS
  F      IFMVBD
  F      IFCHAR
  F      IFVSHT
  F      IFTLOG
  T      IFDIRS
  F      IFPTBG
  F      IFCOMP
  2.00000       2.00000       -1.00000       -1.00000     XFAC,YFAC,XZERO,YZERO
**MESH DATA** 6 lines are X,Y,Z;X,Y,Z. Columns corners 1-4;5-8
  328     3   328            NEL,NDIM,NELV
            ELEMENT     1    [  1A]     GROUP      0
         0.       -0.100000              0.        0.100000
 0.100000              0.       -0.100000              0.
         0.              0.              0.              0.
         0.       -0.100000              0.        0.100000
 0.100000              0.       -0.100000              0.
 0.100000        0.100000        0.100000        0.100000
            ELEMENT     2    [  1B]     GROUP      0
         0.              0.       -0.200000       -0.100000
 0.100000        0.200000              0.              0.
         0.              0.              0.              0.
         0.              0.       -0.200000       -0.100000
 0.100000        0.200000              0.              0.
 0.100000        0.100000        0.100000        0.100000
            ELEMENT     3    [  1C]     GROUP      0
-0.100000       -0.200000              0.              0.
         0.              0.       -0.200000       -0.100000
         0.              0.              0.              0.
-0.100000       -0.200000              0.              0.
         0.              0.       -0.200000       -0.100000
 0.100000        0.100000        0.100000        0.100000
            ELEMENT     4    [  1D]     GROUP      0
         0.              0.        0.200000        0.100000
-0.100000       -0.200000              0.              0.
         0.              0.              0.              0.
         0.              0.        0.200000        0.100000
-0.100000       -0.200000              0.              0.
 0.100000        0.100000        0.100000        0.100000


            ELEMENT  328    [  3 ]     GROUP      0

6326  0.490000       0.951364E-06 -0.959122E-12        0.              0.      C
8326 -0.481000       0.136472E-05 -0.149031E-07        0.              0.      C
2327  0.490000       0.347030E-06 -0.347030E-06        0.              0.      C
4327 -0.481000       0.498403E-06 -0.498403E-06        0.              0.      C
6327  0.490000       0.347030E-06 -0.347030E-06        0.              0.      C
8327 -0.481000       0.498403E-06 -0.498403E-06        0.              0.
```

```
 E  327  1   326.000        3.00000          0.              0.              0.
 E  327  2   33.0000        4.00000          0.              0.              0.
 E  327  3   320.000        1.00000          0.              0.              0.
 E  327  4   54.0000        2.00000          0.              0.              0.
 v  327  5   3.00000        1039.00          0.              0.              0.
 ux=1.7*(1-sin(1.57*x/6.42))
 uy=0
 uz=0
 v  327  6   3.00000        1042.00          0.              0.              0.
 ux=1.7*(1-sin(1.57*x/6.42))
 uy=0
 uz=0
 E  328  1   113.000        4.00000          0.              0.              0.
 E  328  2   156.000        4.00000          0.              0.              0.
 E  328  3   155.000        4.00000          0.              0.              0.
 E  328  4   111.000        4.00000          0.              0.              0.
    328  5        0.             0.          0.              0.              0.
 E  328  6   90.0000        5.00000          0.              0.              0.
  ***** NO THERMAL BOUNDARY CONDITIONS *****
   1 PRESOLVE/RESTART OPTIONS  *****
PRESOLVE
   7          INITIAL CONDITIONS *****
        ux=1.7*(1-sin(3.14*x/6.42))
        uy=0
        uz=0
C Default
C Default
C Default
C Default
  ***** DRIVE FORCE DATA ***** BODY FORCE, FLOW, Q
   4                  Lines of Drive force data follow
        ffx=51000
        ffy=0
        ffz=0
C
  ***** Variable Property Data ***** Overrrides Parameter data.
   1 Lines follow.
   0 PACKETS OF DATA FOLLOW
  ***** HISTORY AND INTEGRAL DATA *****
   6   POINTS.  Hcode, I,J,H,IEL
UVWP      H      2    3    4  303
UVWP      H      5    5    3  110
UVWP      H      5    5    2  110
UVWP      H      3    3    1  162
UVWP      H      2    3    1  162
UVWP      H      5    3    5   24
  ***** OUTPUT FIELD SPECIFICATION *****
   6 SPECIFICATIONS FOLLOW
   F       COORDINATES
   T       VELOCITY
   T       PRESSURE
   F       TEMPERATURE
   F       TEMPERATURE GRADIENT
   0       PASSIVE SCALARS OUTPUTS
  ***** OBJECT SPECIFICATION *****
     0 Surface Objects
     0 Volume  Objects
     0 Edge    Objects
     0 Point   Objects
  ***** CURVED SIDE FORTRAN FUNCTIONS *****
     0 Lines of Fortran
```

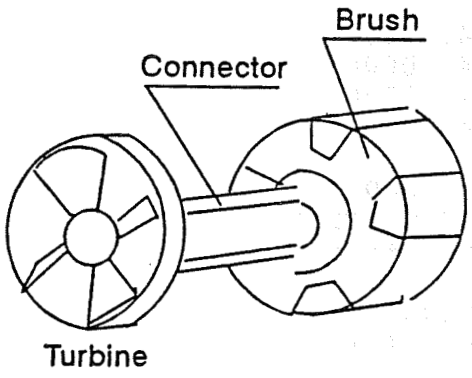
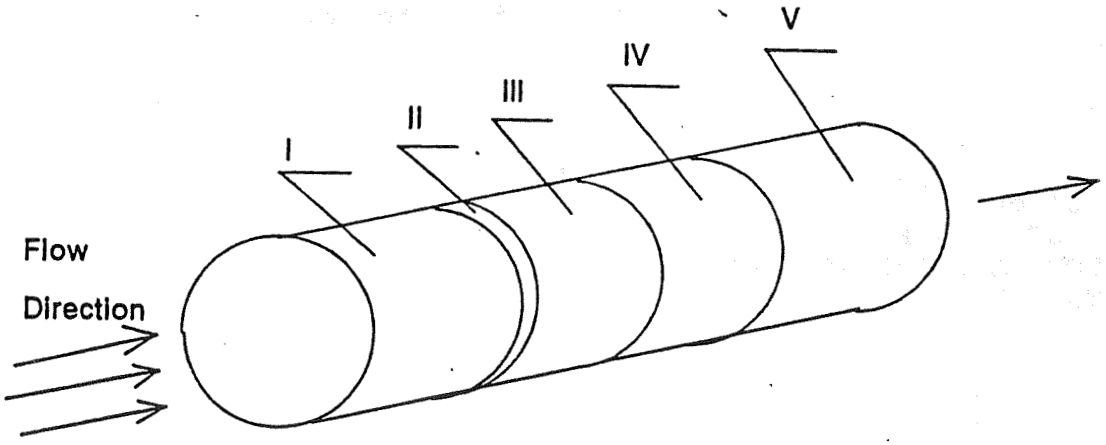


Figure 1. A Sketch of the turbine brush



I:Entrance Region, II:Turbine Section
III:Connector, IV:Brush Section
V:Exit Section of the Flow

Figure 2.Discretization of flow domain

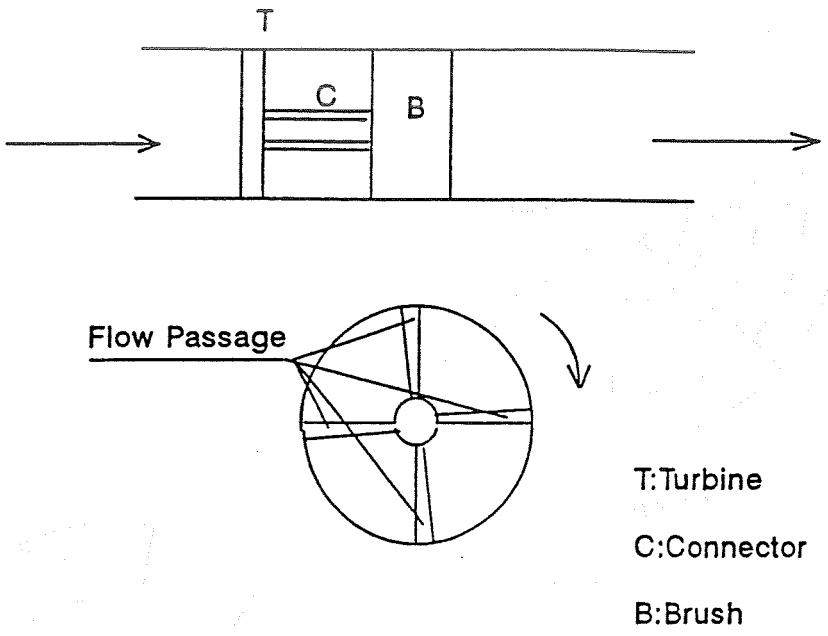


Figure 3. Current model for the turbine

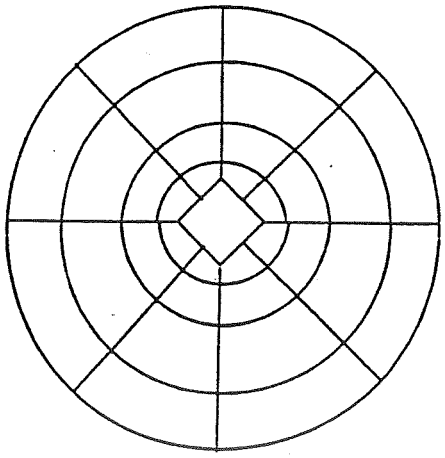


Figure 4. Discretization of Entrance Section

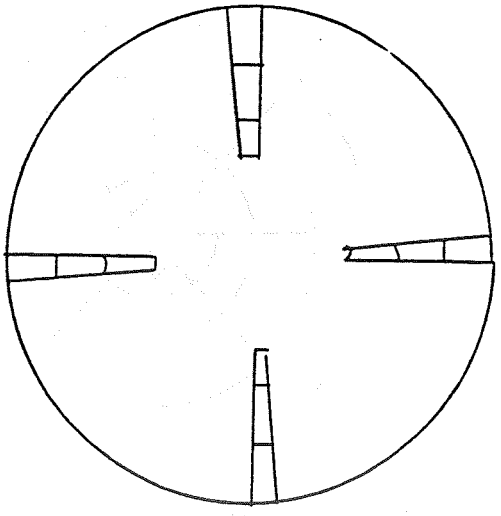


Figure 5. Discretization of Blade Section

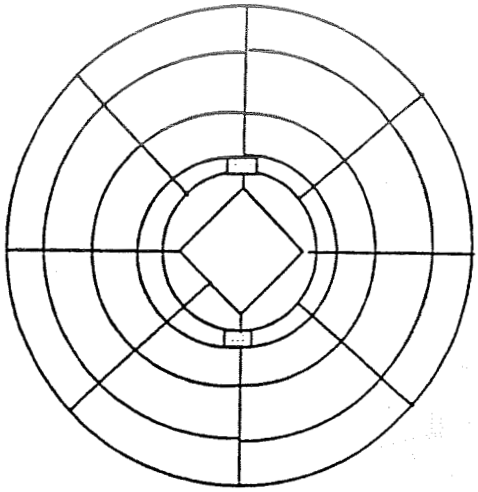


Figure 6. Discretization of Connectors

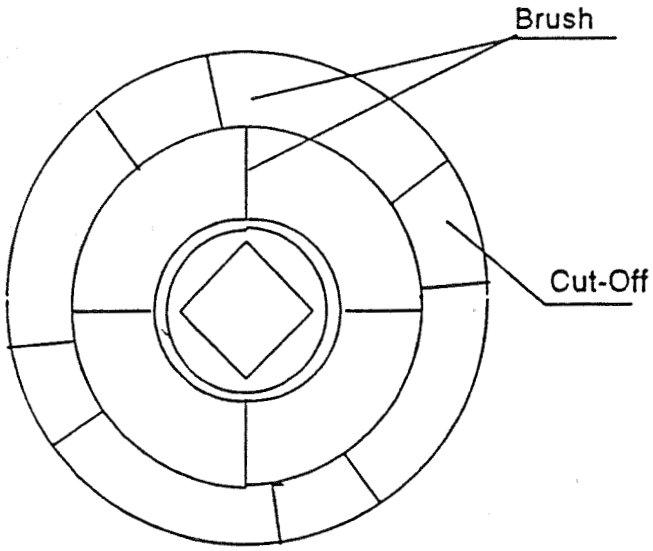


Figure 7. Discretization of Brush

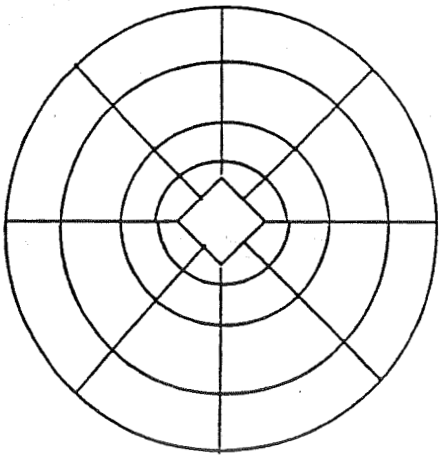
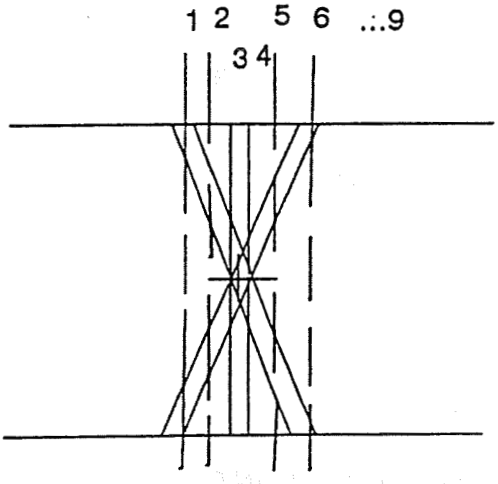
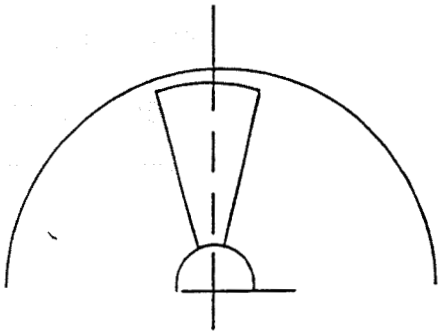


Figure 8. Discretization of Leaving Section



SIDE VIEW

DISTANCE BETWEEN 1 AND 2 = 0.01D



FRONTAL VIEW

Figure 9.Refined Blade Model

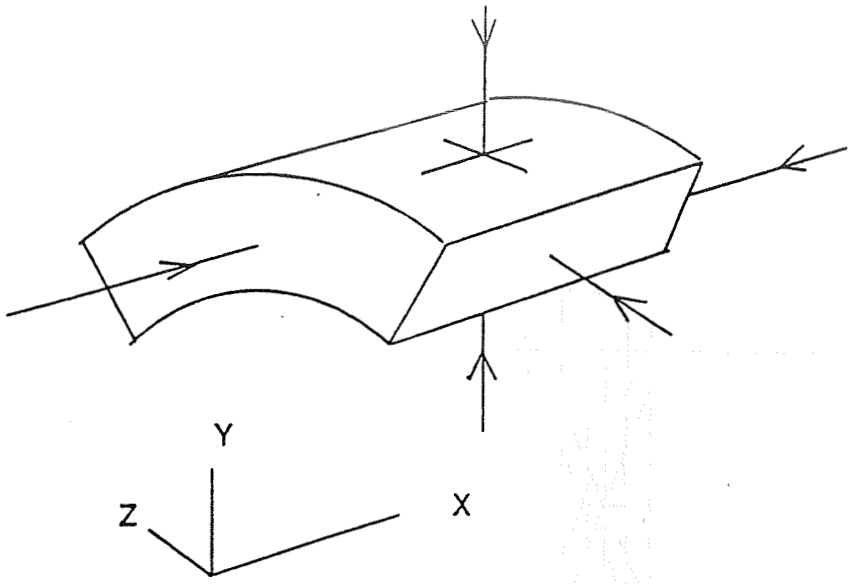
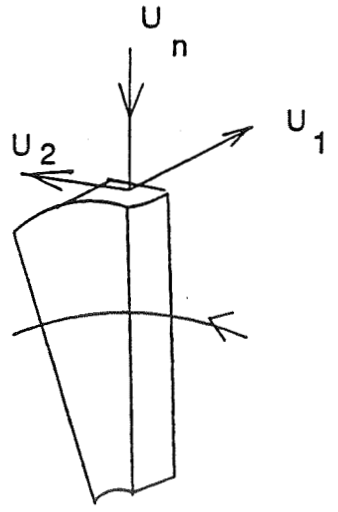


Figure 10.

PRESSURE ON AN INNER ELEMENT

EXPRESSED IN A FUNCTION OR A CONSTANT



$U_n = 0$

$U_1 = 0$

$U_2 = \text{const.}*(y*y+z*z)**(1/2)$

Figure 11.

Velocity Boundary Conditions

Next to a Solid Surface

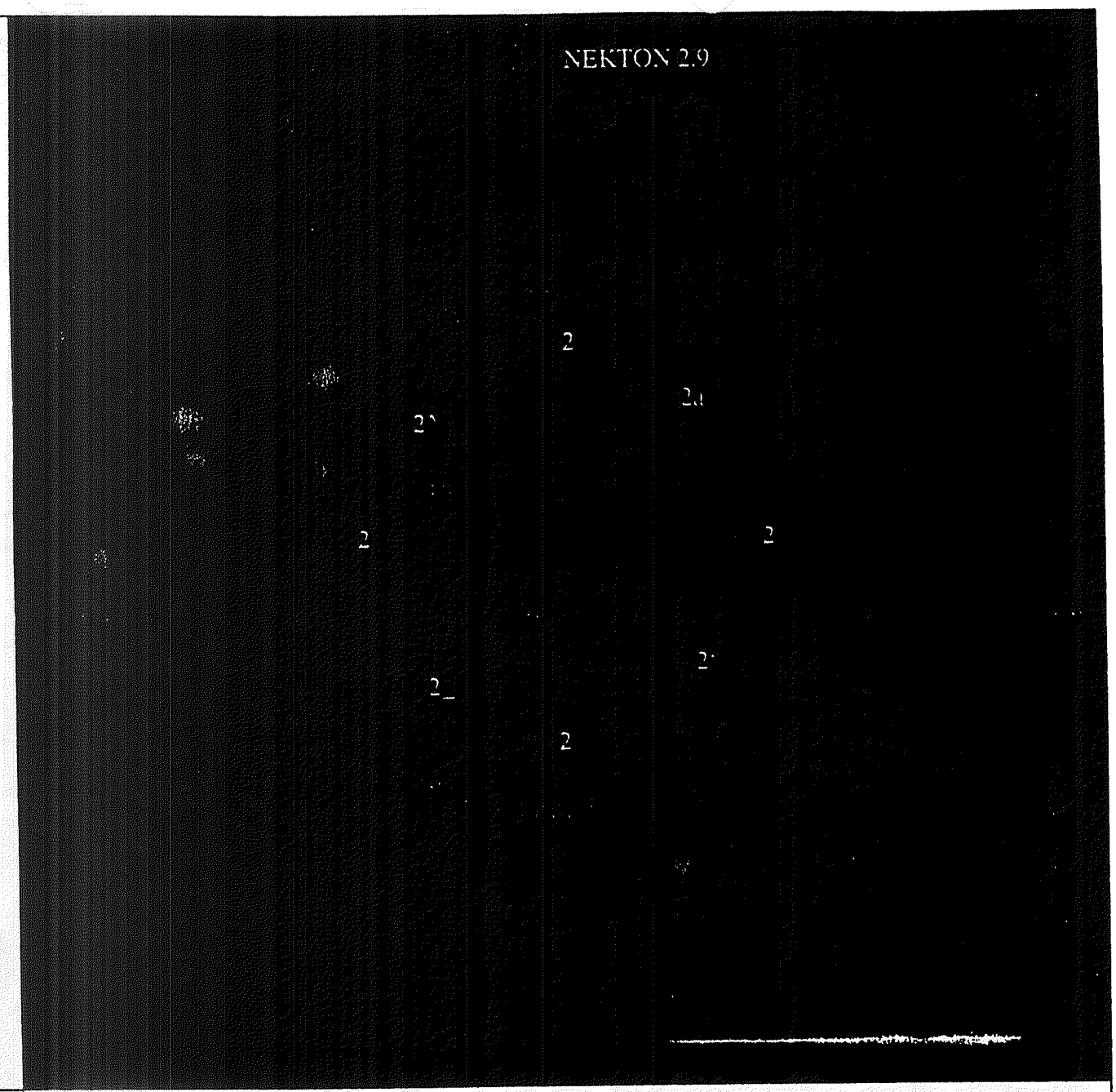


Figure 12a. Front view of the first segment of the blade

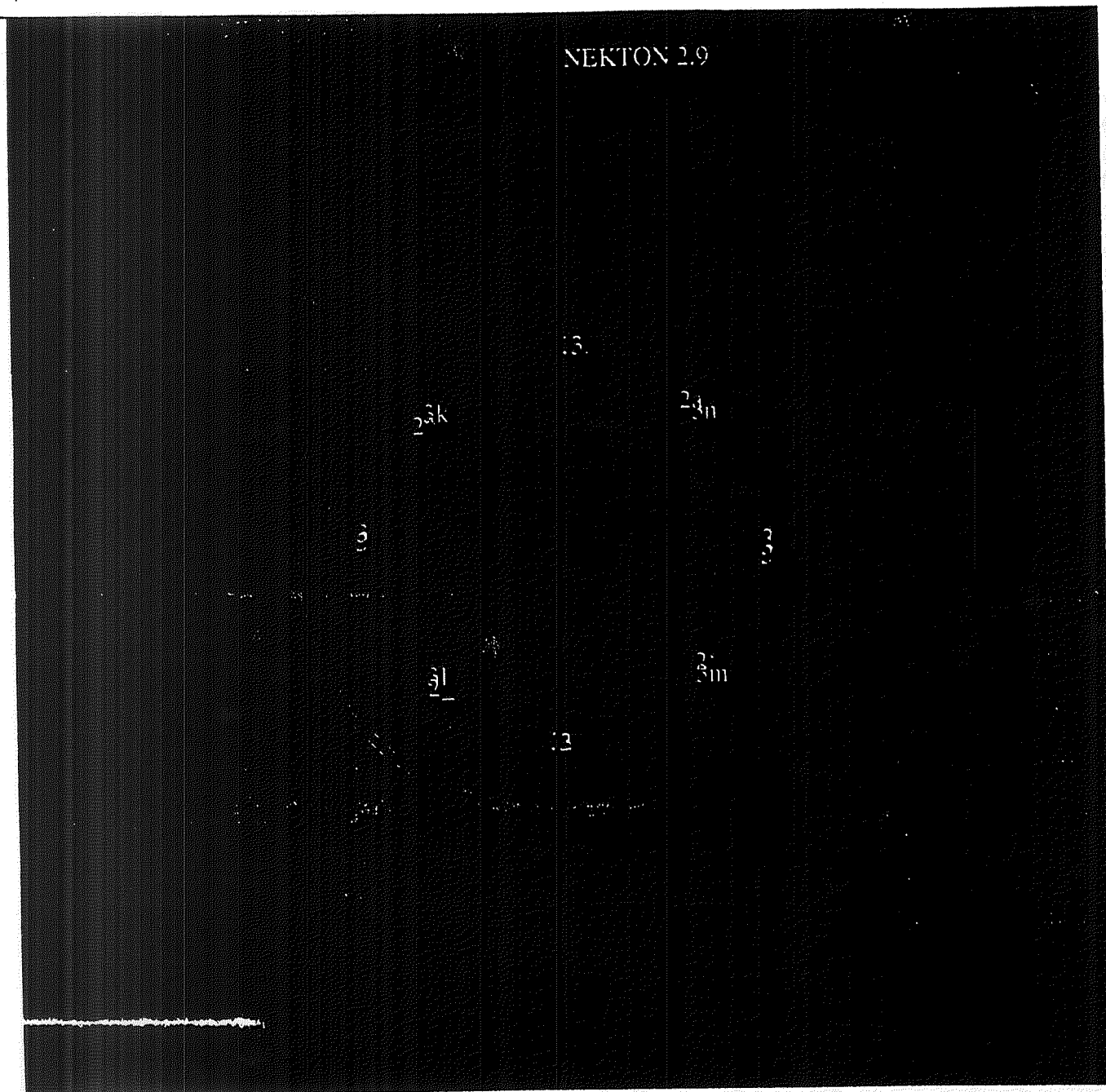
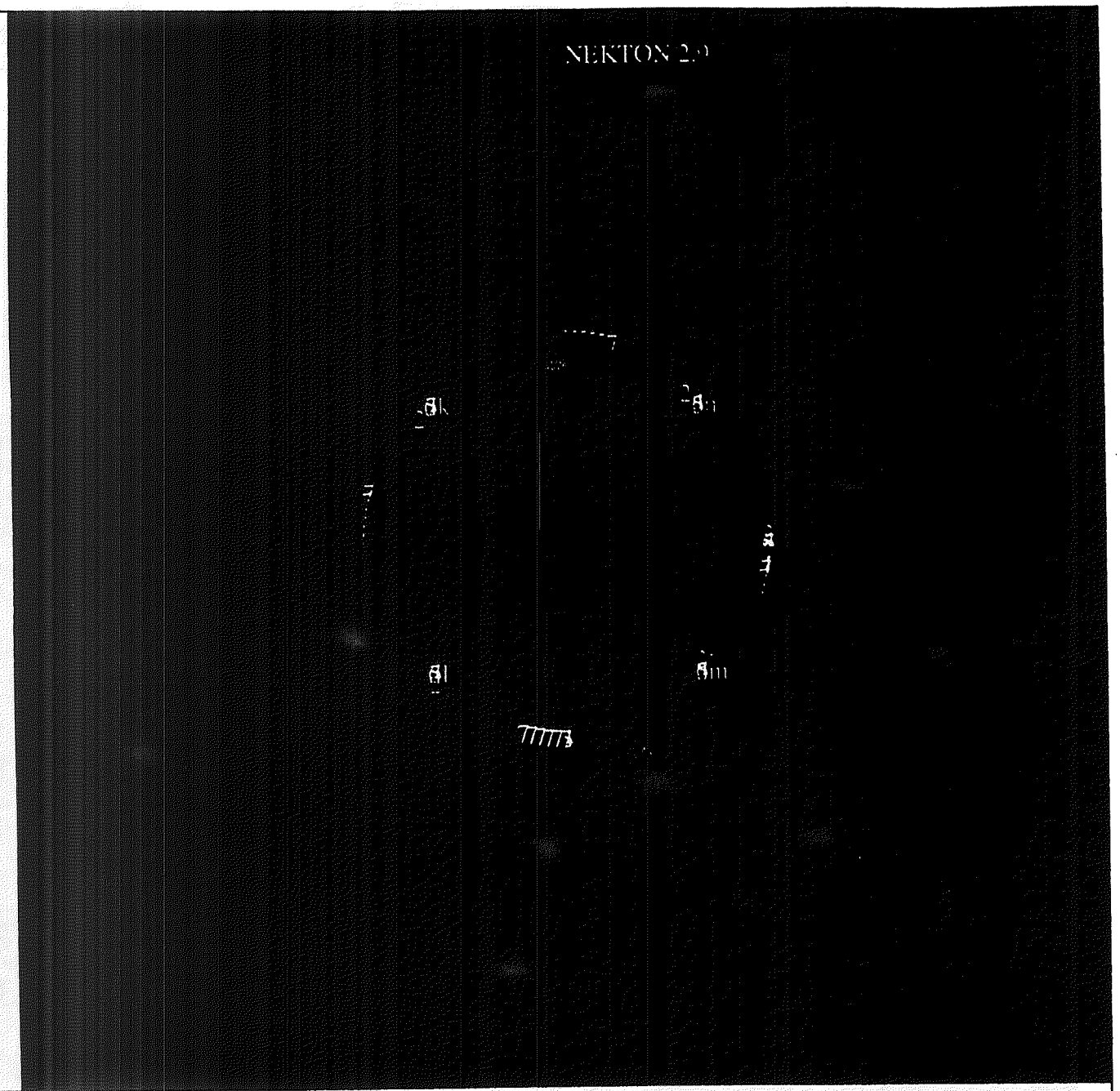


Figure 12b. Front view of the 2nd segment of the blade

Figure 12c. Front view of the 3rd segment of the blade



gure 12d. Front view of the 7th segment the blade

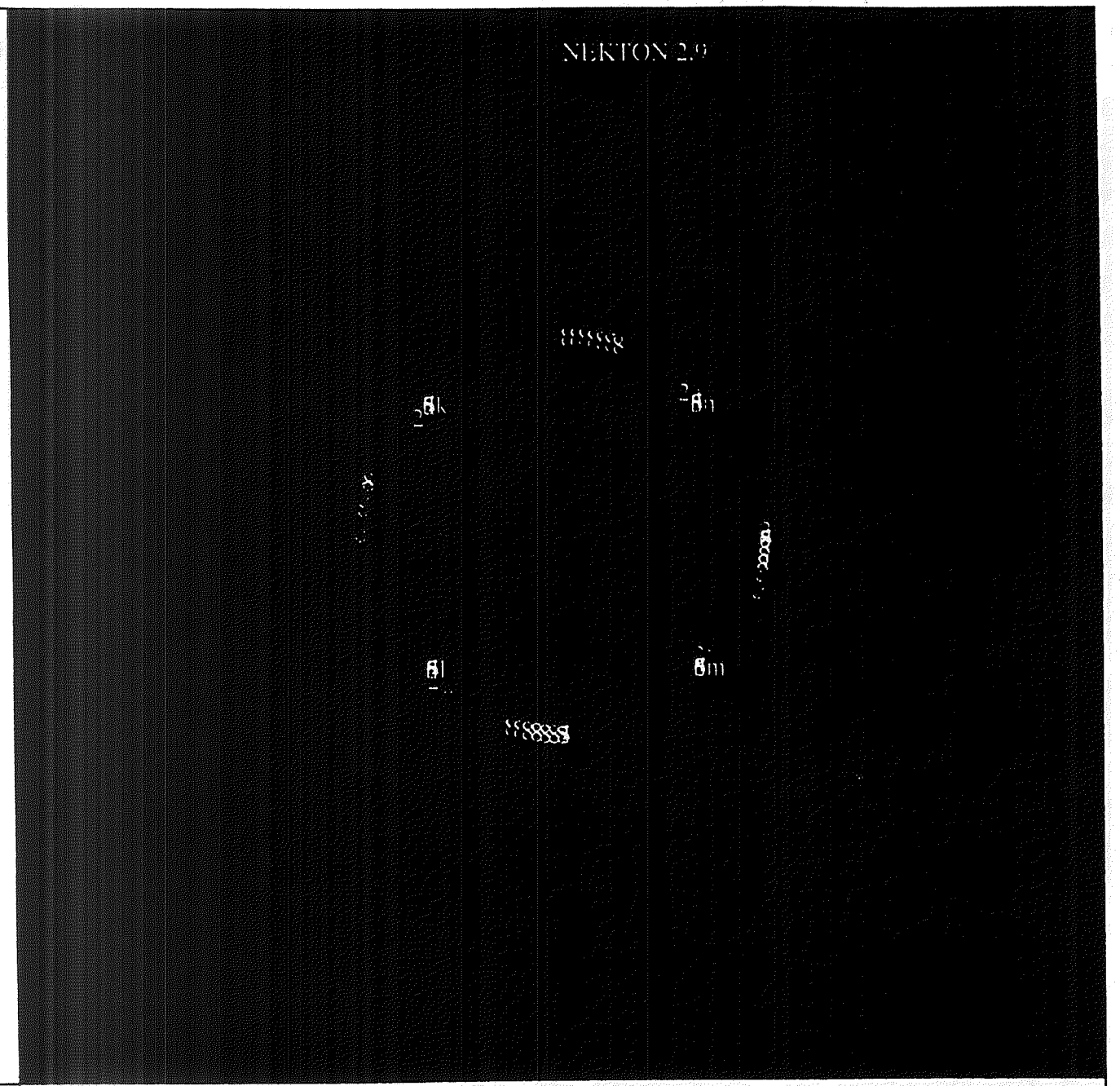
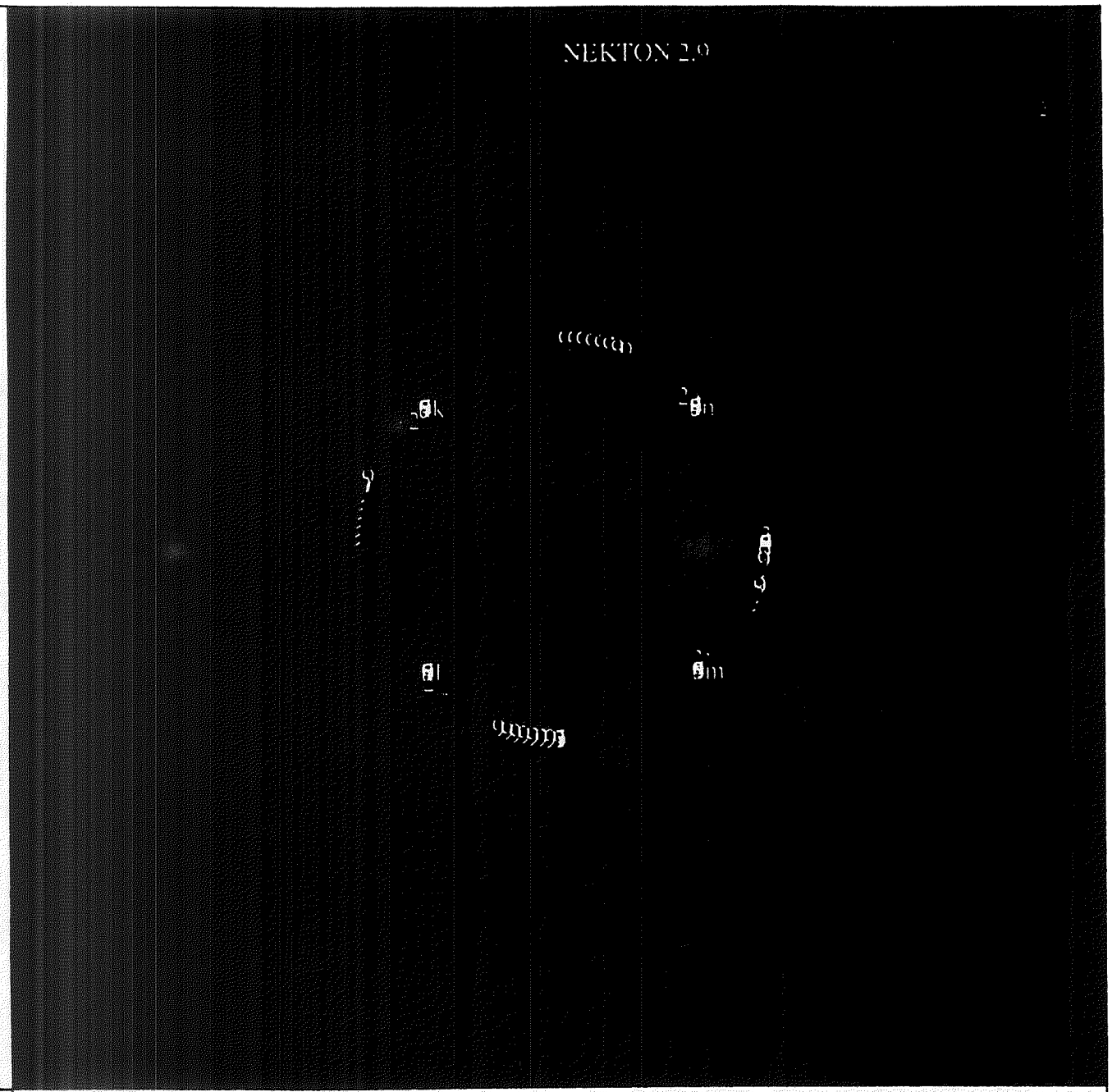


Figure 12e. Front View of the 8th segment of the blade



gure 12f. Front view of the 9th segment the blade

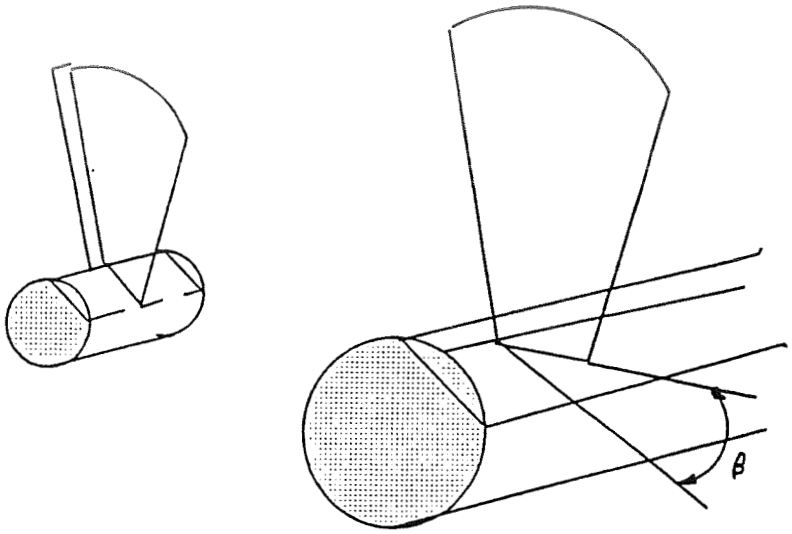


Figure 13. Orientaion of Blade Placement with Respect to the Rotation Axis

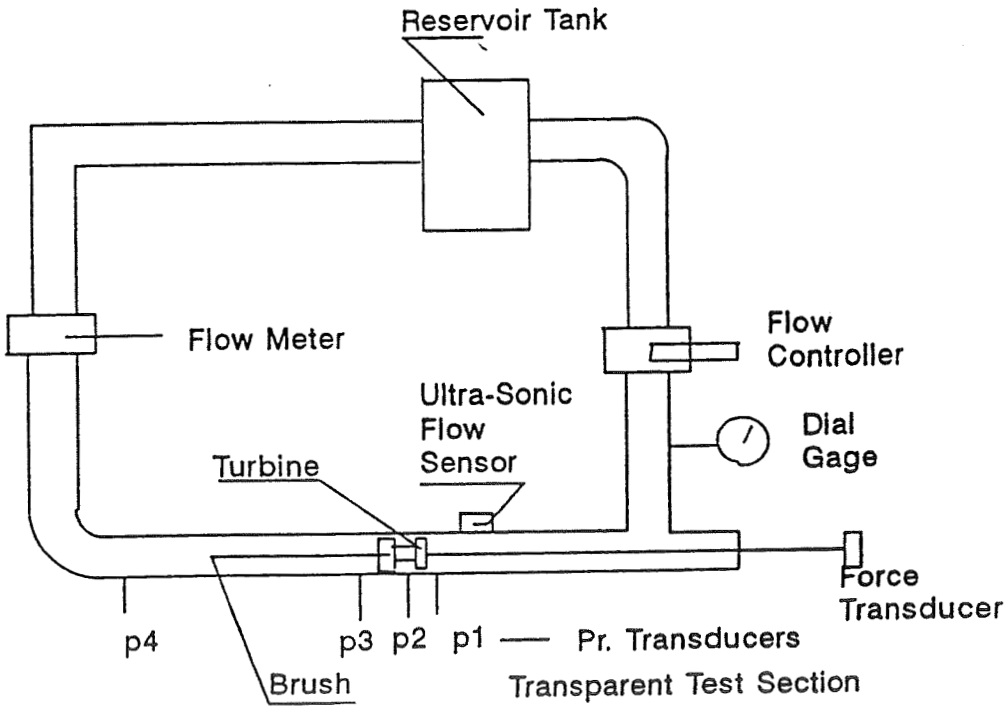


Figure 14. A Closed loop experimental set-up

1995 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

512-80
7752
p. 30

EVALUATION METRICS OF EDUCATIONAL PROGRAMS FOR TEACHERS

Dr. Gwendolyn D. Mitchell
Assistant Professor
Vocational and Adult Education Department
Auburn University
Auburn, Alabama

KSC Colleague - Steve Dutczak
Public Affairs

Contract Number NASA-NGT-60002
Supplement 19

August 23, 1995

PAGE 338 INTENTIONALLY BLANK

ACKNOWLEDGMENTS

This summer experience has been both personally and professionally rewarding. Several people are responsible. To Kari Stiles and Dr. Ray Hosler, thanks for implementing an organized and comprehensive program. To Dr. Steve Dutczak, thank you for providing a comfortable work space and assistance with an interesting project. Thanks Holly for all of your assistance, especially with the computer. To the entire ESB staff, thanks for providing a pleasant work environment.

ABSTRACT

A system for evaluating the teacher programs and services in the Education Services Branch was developed. The primary stakeholder was interested in determining the worth or usefulness of these services to educators. Therefore, two instruments were developed to collect the data. One questionnaire was administered. Data was collected, analyzed and reported. The other questionnaire was pilot tested and will be administered to teachers during the school year.

SUMMARY

Each year, the Education Services Branch (ESB) provides educational services to nearly 1500 educators. Program evaluation is needed to insure that these services continue to meet the needs of teachers.

Two instruments were designed to evaluate the workshops, curriculum updating materials, and Educator's Resource Center (ERC). Questionnaire #1 was designed to gather data on teachers' attitude toward the workshop and ERC. The second questionnaire was developed to gather data on teachers' usage of resource materials and desire for additional training.

The first questionnaire was administered to the Summer Teacher Enhancement Program (STEP) and the University of South Florida (USF) groups on the last day of their workshops. The evaluation data revealed that both groups had a positive attitude toward the workshop and ERC. Both groups also shared several common responses and recommendations for program improvement.

The second questionnaire was field and pilot tested, not administered. There was a low return rate from the educators. Since the pilot test may be representative of the actual study, measures should be taken to induce a high response rate.

TABLE OF CONTENTS

INTRODUCTION AND PROGRAM DESCRIPTION

KSC has many resources available to current and future educators. The Public Affairs Office, Education Services Branch (ESB) provides aerospace workshops, curriculum updating materials, and an Educator's Resource Center to help teachers gain practical aerospace information and experiences. A description of these services is as follows:

Teacher Workshops

Various workshops are offered to help teachers gain a greater understanding of space and aeronautics. The length of these workshops ranges from one to four weeks in the summer and as short as four hours during the school year. Some workshops are held for local teachers. Other workshops include participants from all over the country.

The NASA Education Workshop for Math and Science Teachers (NEWMAST) is targeted toward math, science and technology teachers of grades 7-12. The NASA Education Workshop for Elementary School Teachers (NEWEST) is geared toward teachers of grades 1-6. The Pre-teacher Program is for college students who plan to become teachers. The Summer Teacher Enhancement Program (STEP) is designed for teachers of all subjects and grades.

Other workshops are held for professionals who are interested in being certified to teach. This group includes students from the University of South Florida (USF) and other institutions.

Curriculum Updating Materials

A teacher kit is available to all educators. This kit contains numerous brochures, booklets and audio-visual information on the American space program and the Kennedy Space Center. These publications are designed to help educators update textbook materials as well as provide new aerospace-related information. The purpose of disseminating this information is to assist teachers with incorporating aerospace into their classroom activities and lessons.

Educator's Resource Center

The Educator's Resource Center (ERC) houses a variety of aerospace information. Educators can duplicate all materials, including videotapes, slides, books and other publications, audio cassettes, photographs, laser disks, computer software, and lesson and activity plans. There is no charge for this service.

Each year, ESB provides services to nearly 1500 educators. To insure that these workshops and resource materials continue to meet the needs of teachers, evaluations must be conducted.

EVALUATION MODEL

During the summer, my role was to serve as an external evaluator with the task of creating a system for evaluating the teacher workshops, curriculum updating materials and Educator's Resource Center. Development of an evaluation system requires a model or series of steps. I used Fink & Kosecoff's (1978) model to guide the evaluation. They suggest the following steps:

1. Formulate evaluation questions;
2. Construct evaluation designs;
3. Plan information collection;
4. Collect evaluation information;
5. Conduct information analysis;
6. Report evaluation information; and
7. Manage the evaluation.

Step 1. Formulate Evaluation Questions

The evaluation questions were developed from previous evaluations and meetings with the primary stakeholder, Dr. Steve Dutczak, Chief, ESB. The questions were categorized into four main areas:

Training

1. What are teachers' attitude toward the workshop (topics, activities, instructors and length of time)?
2. What workshop topics did teachers find most useful?
3. What workshop topics did teachers find least useful?
4. Is follow-up training desired?
5. What are teachers' suggestions for improving the workshop?

Materials

1. To what extent are teachers utilizing the resource materials from the teacher kit in their classrooms?
2. What are some examples of #1?
3. Which materials are most useful?
4. Which materials are least useful?
5. What additional materials are needed in the kit?

ERC

1. What services were most helpful?
2. What services were least helpful?
3. Was assistance with using the center readily provided?
4. What are teachers' suggestions for improving the ERC?

Participants' Demographics

1. Gender;
2. Age;
3. Educational background (highest degree obtained, prior knowledge of aerospace before workshop);
4. Grade(s) taught;
5. Subject(s) taught; and
6. School Name and Location.

Step 2. Construct Evaluation Designs

Since the primary stakeholder wanted to know teachers' attitude toward the services after participation or usage, a summative evaluation design was developed for this project. Summative evaluation is conducted to determine the value of a program after completion.

Step 3. Plan Information Collection

Two instruments were developed to gather data. Questionnaire #1, *Aerospace Workshop Evaluation*, was designed to assess participants' attitudes toward the workshop and ERC. Questionnaire #2, *Aerospace Teacher's Kit Evaluation*, was designed to assess teachers': 1) usage of the materials in the kit; and 2) desire for follow-up or additional training. Examples of these questionnaires are located in Appendices A and B.

Step 4. Collect Evaluation Information

Questionnaire #1 was administered to two groups (STEP and USF) on the last day of their workshops. To ensure validity of the instrument, field and pilot tests were conducted before it was administered. The field test consisted of a panel of experts (measurement and content). The two measurement experts were Dr. Andrew Zekeri of Tuskegee University in Tuskegee, Alabama and Dr. Annette Ellis of Morehouse School of Medicine in Atlanta, Georgia. They assessed the instrument to insure that all variables were measurable. Dr. Dutczak and Peggy Ross, Program Management Specialist, were the content experts to insure appropriate use of terminology and clarity of information relative to ESB.

A four-point Likert scale ranging from strongly disagree to strongly agree was used to determine attitudes toward the workshop and materials. Open-ended questions were included to gain more insight into the scaled responses.

The test-retest method (Carmines and Zeller, 1979) was utilized to assess reliability of the instrument for the pilot test. Five STEP participants were asked to serve as the pilot group. They assessed the instrument for overall clarity, length and wording. They were given the first questionnaire and cover letter (Appendix C) on July 17, 1995. One week later, July 24, the pilot group was given the second questionnaire. The questionnaire was revised based on their comments.

A reliability coefficient of 100% was computed by performing the following steps:

- responses on the first questionnaire were compared with responses on the second questionnaire;
- two columns (alike and different) were created;
- responses that were similar on both instruments were recorded in the alike column;
- dissimilar responses were recorded in the different column;
- responses within one point were considered to be similar;
- the alike column was totaled, then divided by the total number of possible points if all responses were similar, and multiplied by 100.

Questionnaire #2 was field tested using the same panel of experts. The pilot test consisted of mailing the instruments and cover letter (Appendix D) to 37 STEP participants from the 1994 workshop. They were asked to complete the questionnaire, comment on overall clarity, length, and wording and return it in the enclosed envelope by August 18, 1995. Eight questionnaires were returned yielding a 22 percent response rate. An open-ended format was used to gather the data. Therefore, the instrument was deemed valid after revisions.

This questionnaire will be administered to all teachers who participated in workshops during the summer of 1995. Surveys will be mailed approximately three months (November 1995) after workshop completion to allow teachers time to use the materials. A second questionnaire will be mailed approximately three months later (February 1996) to the non-respondents.

Step 5. Conduct Information Analysis

Descriptive statistics were used to analyze the data for the two groups (STEP and USF) who completed Questionnaire #1. Frequency tables and column graphs were constructed to display the data. Table and Figure 1 display data results from the STEP participants.

Table 1. STEP responses to Part I of the *Aerospace Workshop Evaluation.*

	RESPONSE SCALE			
	1 Strongly Disagree	2 Disagree	3 Agree	4 Strongly Agree
STATEMENT	**# responded (%)**	**# responded (%)**	**# responded (%)**	**# responded (%)**
1 Instructors were well prepared for class.	1 (3)		15 (47)	16 (50)
2 Workshop topics were relevant to the subjects that I teach.	1 (3)	1 (3)	12 (38)	18 (56)
3 I received practical ideas for applying the information in my class.	1 (3)		12 (38)	19 (59)
4 Instructors stimulated my interest in aerospace.		2 (6)	8 (25)	22 (69)
5 Between workshops, I had ample time to explore the areas that were of special interest to me.	2 (6)	9 (28)	12 (38)	9 (28)
6 Assistance with using the ERC was readily provided.		2 (6)	7 (22)	23 (72)
7 Materials and services in the ERC were useful to me.		1 (3)	10 (31)	21 (66)
8 I plan to share the new information that I acquired with fellow teachers.		1 (3)	5 (16)	26 (81)
TOTAL	5 (2)	16 (6)	81 (32)	154 (60)

STEP

Thirty-three people participated in the STEP Workshop. One questionnaire was incomplete and deemed unusable; therefore, 32 questionnaires were analyzed. Overall, the STEP group had a positive attitude toward the workshop and ERC. Table 1 (TOTAL columns 3 & 4) indicate that 92 percent of the STEP group agreed and strongly agreed with the statements. Only 8 percent disagreed and strongly disagreed with the statements (TOTAL columns 1 & 2).

Figure 1 shows that at least 50 percent of the STEP participants strongly agreed with seven statements. However, statement 5 indicated that 28 percent disagreed as well as strongly agreed.

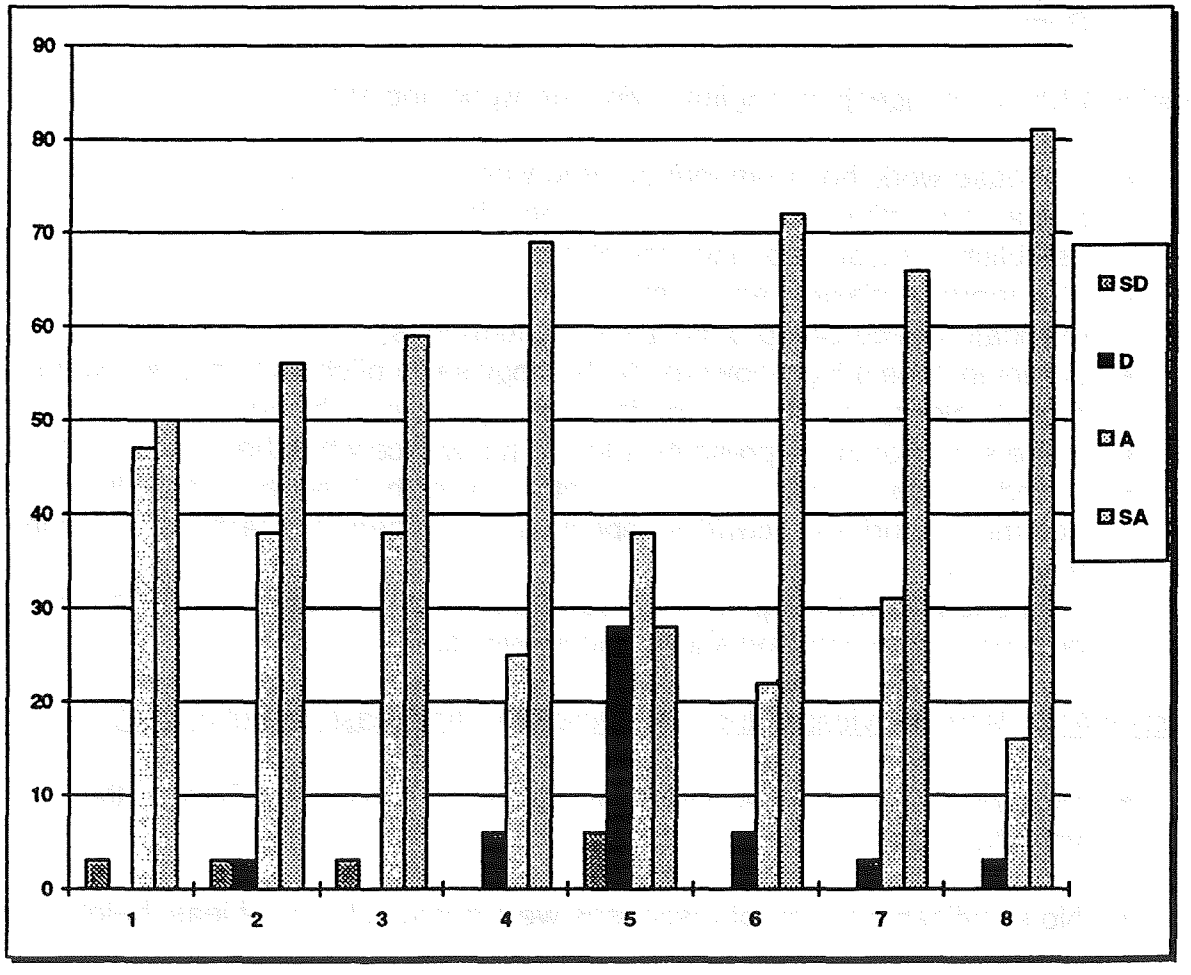


Figure 1. STEP responses by percentage to Part I of the *Aerospace Workshop Evaluation.*

Part II of the *Aerospace Workshop Evaluation* contained four open-ended questions. The responses were as follows:

Question #9: What workshop topics were most and least useful to you?

A total of 90 percent of the STEP group who completed the open-ended questions indicated the most useful workshop topics or activities were:

- job shadowing;
- rocket propulsion;
- KSC tours; and
- technology.

- Sixty percent noted that the least useful topic or activity was the Lego Seminar.

Question #10: My suggestions for improving the workshop are:

- decrease workshop from four to three weeks;
- spread out activities (too much was included in the first week);
- establish hurricane contingency plans;
- plan morning classes and afternoon tours;
- eliminate Space Camp or do it in two afternoons;
- do not include a high-powered technology sales pitch as part of workshops;
- provide clearer details concerning clothing, stipend, housing, etc.;
- make sure workshop presenters know in advance what they are to present;
- schedule Lego demonstration and other topics by teachers' grade level;
- add more hands-on activities especially for elementary teachers, too much time sitting;
- after each workshop, plan KSC tours to show application of material; and
- plan some evening workshops and activities.

Question #11: What services in the ERC were most and least helpful to you?

- Eighty-one percent indicated the most helpful service in ERC was the videotaping.

- No significant amount of responses were noted in terms of least helpful ERC services.

Question 12: My suggestions for improving the ERC are:

- create a time schedule for each teacher to use the equipment to insure that all teachers have an opportunity;
- train staff on using the equipment so that they are able to help teachers;
- provide faster equipment for copying videotapes;
- allow time during the workshop to copy videos; and
- provide Macintosh computer and software.

Part III or Questions 13-20 gathered demographic data on the STEP group. The data revealed the following:

- 81 percent were female;
- 19 percent were male;
- 59 percent have a bachelor's degree;
- 38 percent earned a master's degree;
- 3 percent have specialist degrees;
- 13 percent have less than five years of teaching;
- 31 percent have taught between five and ten years;
- 56 percent have been teaching for more than 10 years;
- all grades (K-12) were taught by the group;
- 3 percent were math teachers;
- 22 percent were science teachers;
- 44 percent taught math and science;
- 31 percent taught other subjects: media, technology, music, agribusiness, computer, general;
- included in the group was a Resource Coordinator;
- 63 percent have not completed other aerospace courses;
- 22 percent have completed 1-3 courses; and
- 15 percent completed more than three courses.

- Everyone did not indicate a school location. Of the participants that responded, most of them were Florida residents. Only three were from Illinois, Colorado and Missouri.

Table 2. USF responses to Part I of the *Aerospace Workshop Evaluation.*

	RESPONSE SCALE			
	1 Strongly Disagree	2 Disagree	3 Agree	4 Strongly Agree
STATEMENT	**# responded (%)**	**# responded (%)**	**# responded (%)**	**# responded (%)**
1 Instructors were well prepared for class.			7 (37)	12 (63)
2 Workshop topics were relevant to the subjects that I teach.			7 (37)	12 (63)
3 I received practical ideas for applying the information in my class.			8 (42)	11 (58)
4 Instructors stimulated my interest in aerospace.			3 (16)	16 (84)
5 Between workshops, I had ample time to explore the areas that were of special interest to me.		8 (42)	3 (16)	8 (42)
6 Assistance with using the ERC was readily provided.	1 (5)	3 (16)	4 (21)	11 (58)
7 Materials and services in the ERC were useful to me.	1 (5)	1 (5)	4 (21)	13 (69)
8 I plan to share the new information that I acquired with fellow teachers.			5 (26)	14 (74)
TOTAL	2 (1)	12 (8)	41 (27)	97 (64)

USF

Table and Figure 2 reveal responses from the USF group. Nineteen questionnaires were completed and analyzed. Overall, the STEP group had a positive attitude toward the workshop and ERC. Table 2 (TOTAL columns 3 & 4) indicate that 91 percent of the USF group strongly agreed and agreed with the statements. Only 9% disagreed and strongly disagreed with the statements (TOTAL columns 1 & 2).

Figure 2 shows that more than 50 percent of the STEP participants strongly agreed with seven statements. However, statement 5 indicated that 28 percent disagreed as well as strongly agreed.

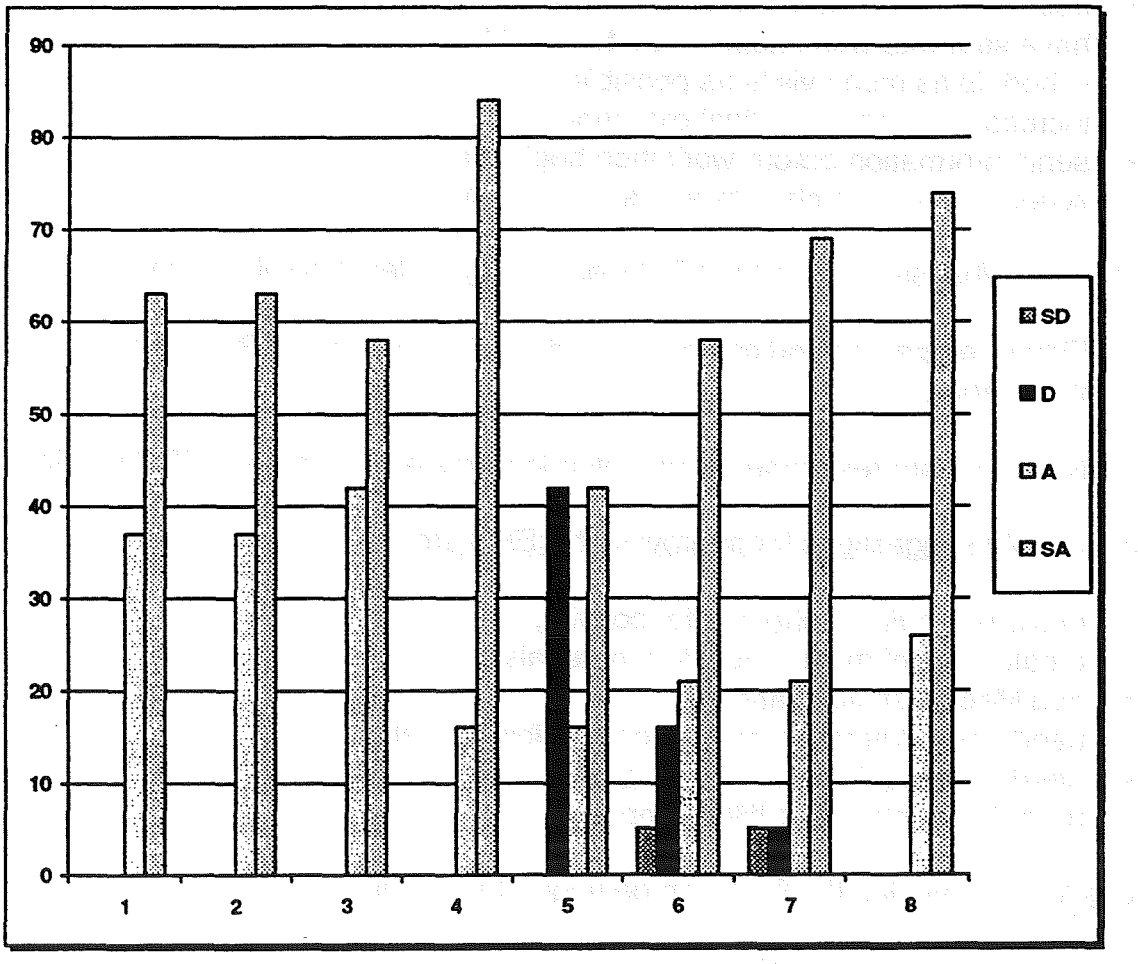


Figure 2. USF responses by percentage to Part I of the *Aerospace Workshop Evaluation.*

USF's responses to the open-ended questions are as follows:

Question #9: What workshop topics were most and least useful to you?

- Eighty percent of the USF group who completed the open-ended questions indicated the most useful workshop topic or activity were the tours.

- No significant amount of responses were noted for least useful workshop.

Question #10: My suggestions for improving the workshop are:

- provide better program coordination;
- add more meaningful hands-on activities;
- have separate workshops for middle and high school teachers;
- schedule as many visits as possible;
- increase time for individual exploring;
- send information before workshop begins; and
- schedule morning classes and afternoon tours.

Question #11: What services in the ERC were most and least helpful to you?

- Fifty-three percent indicated the most helpful service in ERC was the videotaping.

- No significant responses were noted in terms of least helpful ERC services.

Question 12: My suggestions for improving the ERC are:

- designate a time schedule for copying;
- create a list of most requested materials;
- add Macintosh software;
- need new information on IBM compatible formats;
- need more equipment for taping; and
- more information and IBM computers.

Questions 13-20 revealed the following demographic profile:

- 21 percent were female;
- 79 percent were male;
- 47 percent have a bachelor's degree;
- 42 percent have master's degrees;
- 11 percent have doctorates;

- 37 percent have less than five years of teaching;
- 5 percent have taught between five and ten years;
- 58 percent have been teaching for more than 10 years;
- grades 6-12 were taught by the group;
- 21 percent were science teachers;
- 79 were technology education teachers;
- 63 percent have not completed other aerospace courses; and
- 37 percent have completed 1-3 courses.

- All participants were Florida residents.

Step 6. Report Evaluation Information

The results of this evaluation were reported in this written report. Also, a formal presentation was given to all ESB personnel and Public Affairs Directors.

Step 7. Manage the Evaluation

Prior to starting the evaluation, a timetable indicating proposed activities and completion dates was developed (Figure 3). No additional costs were associated with the evaluation. Questionnaires were developed using the office computer. ESB's copy machine was used for duplication. Questionnaire #1 was administered on-site. Questionnaire #2 was mailed through KSC's regular mail system.

EVALUATION TIMELINE

Activity	Completion Date
Research program background	June 26
Review previous evaluations and meet with primary stakeholder	June 26
Compile list of evaluation questions	June 26
Design two instruments: 1. Aerospace Workshop Evaluation 2. Aerospace Teacher's Kit Evaluation	July 7
Field test instruments	July 12
Pilot test and revise instrument #1	July 31
Administer instrument #1	August 4
Pilot test and revise instrument #2	August 18
Present results (oral presentation)	August 23
Formulate and submit written report	August 24

Figure 3. Timetable for managing the evaluation

CONCLUSION AND RECOMMENDATIONS

The data from the *Aerospace Workshop Evaluation* clearly showed that STEP and USF had favorable opinions of the workshop and ERC. Both groups shared these common responses and recommendations for improvement:

- duplicating videotapes was the most helpful ERC service;
- create a time schedule to insure that everyone has an opportunity to duplicate tapes;
- schedule morning classes and afternoon tours (afternoon lectures are not productive);
- provide more hands-on activities;
- add more Macintosh computers and software in the ERC; and
- schedule more time for teachers to explore their individual interests during the workshop.

Since both STEP and USF indicated these responses, it is recommended that the ESB staff consider these suggestions for implementation during future workshops.

The return rate on the pilot test for the *Aerospace Teacher's Kit Evaluation* was low. A low return rate may occur when the actual questionnaires are mailed during the school year. To increase the number of completed and returned questionnaires, several items are recommended:

- mail the questionnaires to educators approximately three months after their summer workshop to allow them time to use the materials in class;
- include a stamped and addressed envelope with the questionnaire;
- allow them ample time for them to respond;
- send another questionnaire in February to the non-respondents;
- provide an incentive for them to complete and return the questionnaires by the due date;
- mail questionnaires to the school principal;
- give questionnaires to Spacemobile representatives to take with them to the schools and distribute to teachers; and
- keep questionnaires in the ERC and distribute to teachers.

REFERENCES

Carmines, E. and Zeller, R. (1979). *Reliability and validity assessment.* Sage Publications: Beverly Hills.

Fink, A. and Kosecoff, J. (1978). *An evaluation primer.* Sage Publications: Beverly Hills.

National Aeronautics and Space Administration, John F. Kennedy Space Center. (1995). *NASA facts: NASA and contractor educational programs at the John F. Kennedy Space Center* (KSC Release No. 138-95). Kennedy Space Center, FL: Author.

APPENDIX A

Aerospace Workshop Evaluation

National Aeronautics and
Space Administration
John F. Kennedy Space Center
Kennedy Space Center, FL 32899

AEROSPACE WORKSHOP EVALUATION

Your opinion is extremely important to us. Please take a few minutes to complete the survey to let us know your feelings toward this workshop.

Part I. Please respond to each statement using the following key (circle your answer):

Key	
1=Strongly Disagree	3=Agree
2=Disagree	4=Strongly Agree

		Strongly Disagree			Strongly Agree
1.	Instructors were well prepared for class.	1	2	3	4
2.	Workshop topics were relevant to the subjects that I teach.	1	2	3	4
3.	I received practical ideas for applying the information in my class.	1	2	3	4
4.	Instructors stimulated my interest in aerospace.	1	2	3	4
5.	Between workshops, I had ample time to explore the areas that were of special interest to me.	1	2	3	4
6.	Assistance with using the Educators Resource Center (ERC) was readily provided.	1	2	3	4
7.	Materials and services in the ERC were useful to me.	1	2	3	4
8.	I plan to share the new information that I acquired with fellow teachers.	1	2	3	4

Part II. Please answer the following questions:

9. What workshop topics were:
   - most useful to you?

   - least useful to you?

10. My suggestions for improving the workshop are:

11. What services in the ERC were:
    - most helpful to you?

    - least helpful to you?

12. My suggestions for improving the ERC are:

Part III. Please circle the appropriate information or fill in the blanks (where applicable):

13. Gender: 1. Female 2. Male

14. Highest degree obtained: 1. Bachelor's 2. Master's 3. Specialist 4. Doctorate

15. Number of years teaching: 1. Less than Five 2. Five-Ten 3. More than Ten

16. Grade level(s) you teach: K 1 2 3 4 5 6 7 8 9 10 11 12

17. Subject(s) you teach: 1. Mathematics 2. Science 3. Other _____

18. Number of courses you've completed related to aerospace (excluding this workshop):

    1. Zero 2. One-Three 3. More than Three

19. Name (optional): _____

20. School name & location: _____

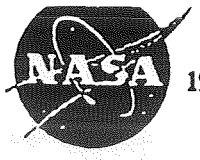
    _____

THANK YOU!

APPENDIX B

Aerospace Teacher's Kit Evaluation

National Aeronautics and
Space Administration
John F. Kennedy Space Center
Kennedy Space Center, FL 32899



AEROSPACE TEACHER'S KIT EVALUATION

*Part I. **Listed below are titles of most of the booklets and brochures in your kit. Please answer the questions below:***

1. America's Spaceport
2. Astronaut Fact Book
3. Astronaut Selection and Training
4. Countdown! NASA Launch Vehicles and Facilities
5. Living and Working on the New Frontier
6. Materials Processing in Space
7. Questions and Answers About Aeronautics and Space
8. Spacelab
9. Space Link
10. Space Shuttle
11. Space Shuttle Mission Chronology STS-1 - STS-61 1981-1993
12. Space Shuttle Mission Summary The First Decade: 1981-1990
13. Spinoffs
14. Teacher Resource Center Network
15. The Early Years: Mercury to Apollo-Soyuz
16. Video Catalog
17. Wardrobe for Space

1. Which of the publications have you used to create classroom activities (indicate numbers only)?

2. Indicate the activities you've developed (if possible, provide sample lesson plans or continue details on back).

3. Do you need a training workshop to help you fully incorporate these materials into your class lessons? _____YES _____NO

4. What other workshops would help you to provide comprehensive lessons related to space and aeronautics?

5. What additional information would you like to see in the kit?

Part II. Please circle the appropriate information or fill in the blanks (where applicable):

6. Name: _____

7. School name & location: _____

_____

8. Gender: 1. Female 2. Male

9. Highest degree obtained: 1. Bachelor's 2. Master's 3. Specialist 4. Doctorate

10. Number of years teaching: 1. Less than Five 2. Five-Ten 3. More than Ten

11. Grade level(s) you teach: K 1 2 3 4 5 6 7 8 9 10 11 12

12. Subject(s) you teach: 1. Mathematics 2. Science 3. Other _____

Thanks for your assistance!

Use this space to provide additional information

APPENDIX C

Cover Letter accompanying the Aerospace Workshop Evaluation
(pilot test)

National Aeronautics and
Space Administration
John F. Kennedy Space Center
Kennedy Space Center, FL 32899

July 13, 1995

Dear Teacher:

In the near future, the Education Services Branch will use a new type of questionnaire for workshop evaluations. You have been selected to be a part of the pilot group to test this new questionnaire. ***Please complete the survey and comment on the following:***

1. *Length of time to complete _____ minutes*

2. *Wording (too difficult, too simple, just right* _____

_____

3. *Length of survey (too long, too short, just right)* _____

_____

4. *Overall clarity (easy to understand, not clear, just right)* _____

_____

Since the test-retest method is being used to insure the reliability of this questionnaire, you will be given another copy to complete in one week. The two sets of responses will be compared and a reliability coefficient will be calculated. Thanks for your assistance!

Sincerely,

Steve Dutczak, Chief
Education Services Branch

APPENDIX D

Cover letter accompanying the Aerospace Teacher's Kit Evaluation

National Aeronautics and
Space Administration
John F. Kennedy Space Center
Kennedy Space Center, FL 32899

August 7, 1995

Reply to Attn of: PA-ESB

Dear Teacher:

Last summer you received an Aerospace Teacher's Kit from the Education Services Branch at the Kennedy Space Center. In the near future, we will be conducting a survey to determine the extent to which teachers incorporate this information into their classroom activities. You have been selected to be a part of the pilot group to test the new questionnaire. **If you have used any of the information from these materials to teach lessons related to space and aeronautics, please take a few minutes to complete the survey and comment on the following:**

1. *Length of time to complete the survey: _____ minutes*

2. *Wording (too difficult, too simple, ok):*

3. *Length of survey (too long, too short, ok):*

4. *Overall clarity (easy to understand, not clear, ok):*

Please return the completed survey and your comments by Friday, August 18, 1995. It is crucial that we receive your response by this date. A return envelope is enclosed for your convenience. We greatly appreciate your assistance!

Sincerely,

Steve Dutczak, Chief
Education Services Branch

Enclosure

1995 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

513-35
7753
V. 26
(26)

SENSING STRATEGIES FOR TOXIC VAPOR DETECTION

Dr. Horacio A. Mottola
Regents Professor
Department of Chemistry
Oklahoma State University
Stillwater, Oklahoma

KSC Colleague - Dale Lueck
Instrumentation and Hazardous Gas Monitoring

Contract Number NASA-NGT-60002
Supplement 19

August 9, 1995

TABLE OF CONTENT

ACKNOWLEDGEMENTS

I would like to express my appreciation for being selected as a 1995 NASA/ASEE Summer Fellow at Kennedy Space Center. I am particularly indebted to all the people in the Toxic Vapor Detection/Contamination Monitoring Laboratory (LC 34) for their friendship and their constant facilitation of my work. Thanks are extended to my NASA colleague Dale Lueck (NASA Instrumentation Section) for providing me with the opportunity to participate in this project. Finally, I wish to acknowledge the administrative support of Mr. Gregg Buckingham and Dr. Ray Hosler, and the always friendly help received of Mrs. Kari Stiles. Matt Russell, a senior at Cocoa Beach High School, helped in some phases of this work and provided the opportunity and the challenge of having to explain at his level why things did or did not work.

ABSTRACT

This work was motivated by the recommendation of the American Conference of Governmental Industrial Hygienists (ACGIH) that threshold limits for hydrazine, H_2N-NH_2, in air be lowered from 100 to 10 parts-per-billion (ppb) concentration levels. Hydrazine is one of the high-energy propellants used in large volumes in Space Shuttle, Titan, payloads, and other aerospace operations. Since analytical methods presently available for hydrazine detection and/or determination do not satisfy such low levels of detection, the ultimate goal of this research is the development and characterization of a portable and compact chemical sensor ideally capable to detect (in real time) 1 ppb of hydrazine, continuously and reversibly. The laboratory prototype developed as part of this project is comprised of: (1) a reactor part in which H_2N-NH_2 reacts, generating chemiluminescence emission, with tris[2,2'-bipyridine]ruthenium(III), which is immobilized on an ion-exchange polymeric material of a perfluorinated hydrocarbon containing sulfonate groups as exchange centers (Nafion), (2) an electrochemical three-electrode cell posed at a potential at which the immobilized ruthenium complex could be reoxidized to the III-oxidation state (as to provide reversible and continuous detection), and (3) a low power consumption photomultiplier tube to collect and quantitatively integrate the emitted photons with the help of auxiliary electronics and readout device.

SUMMARY

The scope of this project was to test the feasibility of hydrazine detection with a sensor designed to integrate, in a single unit, chemiluminescence detection and electrochemical regeneration of the main reagent. The unit designed and built as part of this research incorporated the main reagent (a ruthenium complex) immobilized on an ion exchange-polymer (perfluorosulfonic acid). Cyclic voltammetry (an electroanalytical technique based on applying a programmed potential which varies, at a given rate and in a cyclic form, between an initial and a switching potential) was used to characterize the polymeric films with the immobilized main reagent. Limitations of the approach, avenues for improvement, as well as alternatives for further sensor development are presented in the text of this final report.

LIST OF ILLUSTRATIONS

I. INTRODUCTION

1.1. THE NEED FOR A HYDRAZINE SENSOR

Hydrazine is one of the high-energy propellants used in large volumes in Space Shuttles, Titan, payloads, and other space operations. Hydrazine has been identified as a carcinogen or potential carcinogen. It may cause lung damage and upper respiratory tract irritation if inhaled, and relatively high concentrations of hydrazine may be fatal. The acute oral LD_{50} in rats, ingested as hydrazine sulfate, is 50-400 $mg.Kg^{-1}$ [1]. Alergic skin irritations are known to be caused by skin contact with hydrazine or hydrazine salts.

The American Conference of Governmental Industrial Hygienists (ACGIH) has recommended that threshold limits for hydrazine in air be lowered from 100 to 10 parts-per-billion (ppb) [2]. Compliance with this recommendation requires highly sensitive methods of detection, and present methods in use are not capable of satisfying such requirement.

1.2. HYDRAZINE CHEMILUMINESCENCE REACTION WITH TRIS[2,2'-BIPYRIDINE]RUTHENIUM(III)

Chemiluminescence (emission of light resulting from a chemical reaction) offers very high sensitivity (low limits of detection) and wide dynamic concentration ranges for analytical determinations [3]. In 1966 Hercules and Lytle [4] reported the generation of visible light when tris[2,2'-bipyridine]ruthenium(III), $Ru(bpy)_3^{3+}$, was reduced either by hydrazine or hydroxide ion. Since then, several other papers have appeared in the literature discussing the mechanism and kinetics of this chemiluminescence, as well as the reduction by several aliphatic amines leading to light emission [5]. The chemical structure of the $Ru(bpy)_3^{2+}$ is shown in Figure 1.

The chemiluminescence is probably due to the formation of a short-lived amine radical cation resulting from the reduction of $Ru(bpy)_3^{3+}$ by hydrazine, which further reacts with

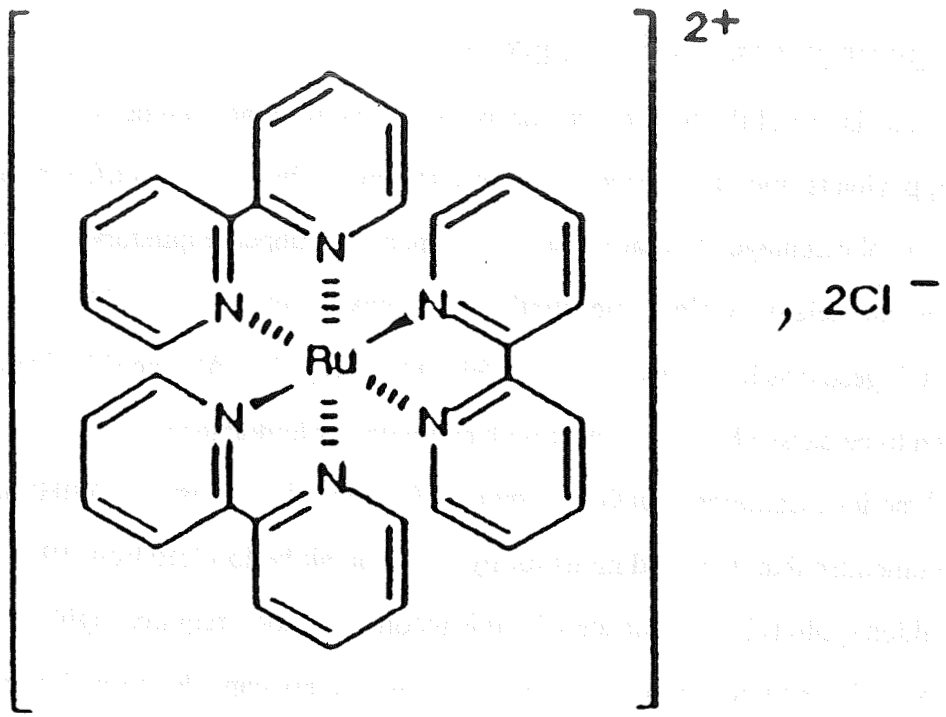


FIGURE 1. Chemical arrangement in the tris[2,2'-bipyridine]ruthenium(II) complex.

$Ru(bpy)_3^{3+}$ or excess $Ru(bpy)_3^{2+}$ to form the excited state of the Ru(II)-complex as shown in the following chemicophysical sequence:

$$Ru(bpy)_3^{3+} + \text{hydrazine} \longrightarrow [Ru(bpy)_2^{2+}]^*$$

$$[Ru(bpy)_3^{2+}]^* \longrightarrow Ru(bpy)_2^{2+} + h\upsilon \ (600\ nm)$$

in which $[Ru(bpy)_3^{2+}]^*$ indicates the excited state of the Ru(II)-complex. The process is very rapid and offers the potential of affording very low limits of detection, hence its selection for this project. Lee and Nieman [6] have recently provided a brief review of mechanistic aspects of the chemiluminescent reactions involving $Ru(bpy)_3^{3+}$ and their analytical applications.

1.3. ELECTROSTATIC IMMOBILIZATION OF TRIS[2,2'-BIPYRIDINE]RUTHENIUM(II)/(III) AND ELECTROCHEMICAL GENERATION OF THE Ru(III) COMPLEX

The orange-red ruthenium(II) complex can be electrostatically immobilized on a cation exchanger in the general form illustrated below:

$$\{(R{\approx}SO_3^-)_n,\ (n/2)\ [Ru(bpy)_3^{2+}]\}$$

in which $(R{\approx}SO_3^-)_n$ represents a typically sulfonated cation-exchange site in a polymeric backbone R.

Nafion, the trade mark for a chemically stable perfluorosulfonic acid polymer (E. I. du Pont de Nemours, Wilmington, DE) commonly used in membrane form for the manufacture of chlorine and caustic soda [7], is effective in storing in insoluble form significant amounts of $Ru(bpy)_3^{2+}$ [8]. The green ruthenium(III) complex was prepared and immobilized on Nafion by a similar route during the course of these studies, and for this reason these two complexes were singled out for use in this work. The direct attachment of the ruthenium(III) complex was made part of these studies in an attempt to circumvent that when the ruthenium(II) complex is the main electroactive species on the surface of the screen, only a fraction of its total amount is converted to the III oxidation state and the electrogenerated $Ru(bpy)_3^{3+}$ is the limiting reagent for chemiluminescence.

Oxidation of the immobilized ruthenium(II) complex is possible chemically (e.g. by reaction with lead dioxide or chlorine gas) or electrochemically at an applied potential of about +1.38 V vs. a silver/silver chloride (Ag/AgCl, 3.0 M NaCl) reference electrode,

according to:

$$Ru(bpy)_3^{2+} \longrightarrow Ru(bpy)_3^{3+} + e^-$$

Oxidation at potentials higher than +1.50 V need to be avoided because of electrolysis of water at platinum surfaces. Electrochemical oxidation of the ruthenium(II) complex affords a clean avenue for the regeneration of the main reagent [i.e., the Ru(III) complex], and can be incorporated into compact sensing units.

Consequently, with the prototype sensor developed during this work, the utilization of electrochemical regeneration of the main reagent at a platinum screen working electrode as part of a three-electrode cell (Figure 2) was included as part of strategies explored.

Figure 3 provides a closer view of the platinum screen working electrode which also serves as receptacle for the immobilized ruthenium redox couple.

1.4. NEED FOR COMPENSATION OF SOLVENT EVAPORATION

When $Ru(bpy)_3^{2+}$ is the immobilized species, continuous contact of the polymeric film of Nafion-immobilized Ru(II/III) complex with the air sample (potentially containing the analyte hydrazine) should result in superficial drying of the film impairing the ionic mobility required for sustainment of the electrochemical regeneration of the Ru(III) centers. Moreover, such a drying can result in flaking of the polymer film, and eventual deterioration of the sensing unit. Nafion is a hydrophilic material and can be rehydrated. The rate of rehydration, however, is not fast and air-drying may prevail between these two processes if a device for compensation for solvent evaporation is not part of the overall

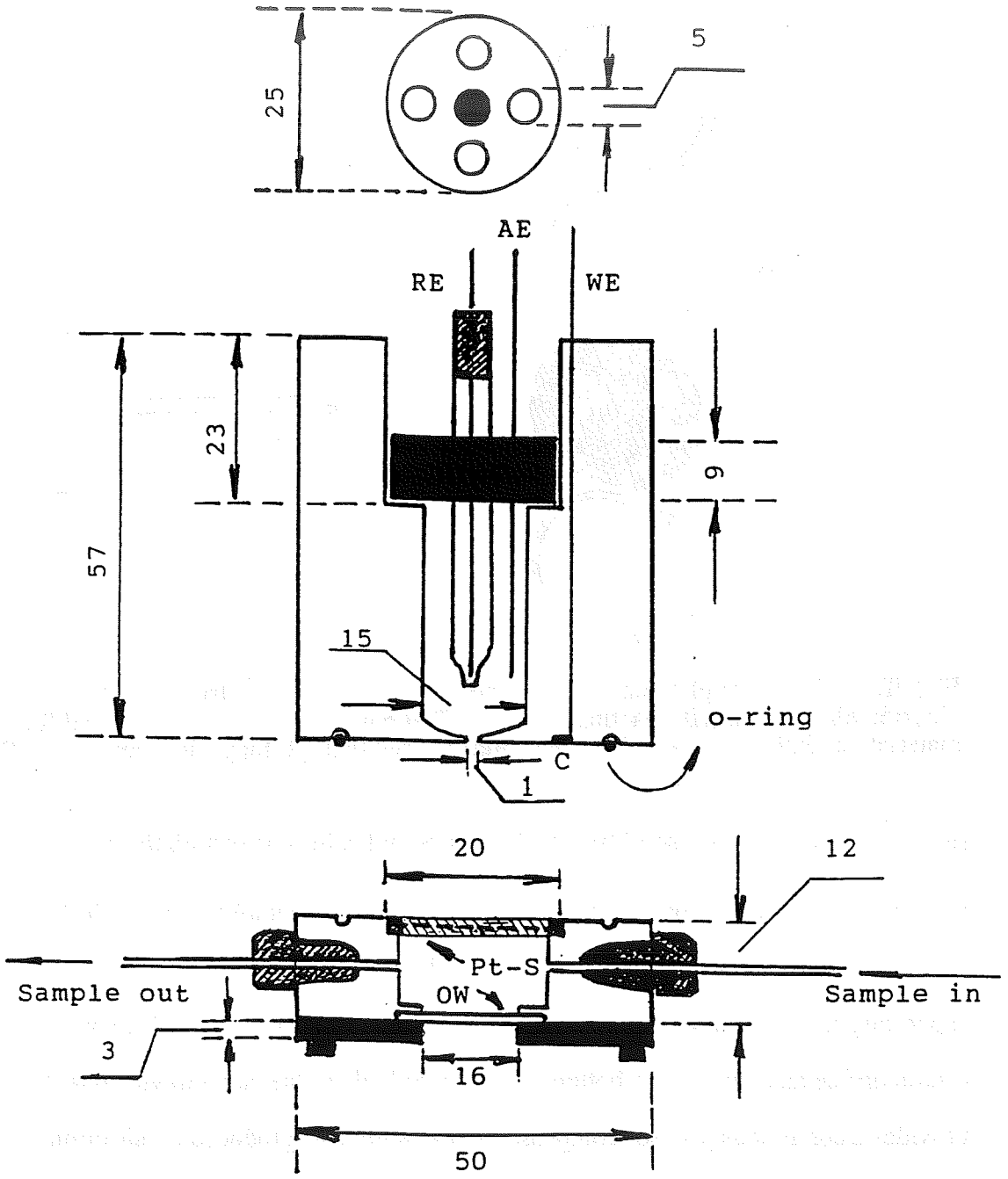


FIGURE 2. Schematic diagram of integrated chemiluminescence reactor chamber, three-electrode reagent regeneration cell, and optical window for photon detection. Measurements in mm. RE: reference electrode (Ag/AgCl/3.0 M NaCl); AE: auxiliary electrode (platinumwire); WE: working electrode connection (working electrode: Pt-S, platinum screen); C: copper contact; OW: optical window (to photomultiplier tube). For further details see narrative (INSTRUMENTATION under MATERIALS AND METHODS).

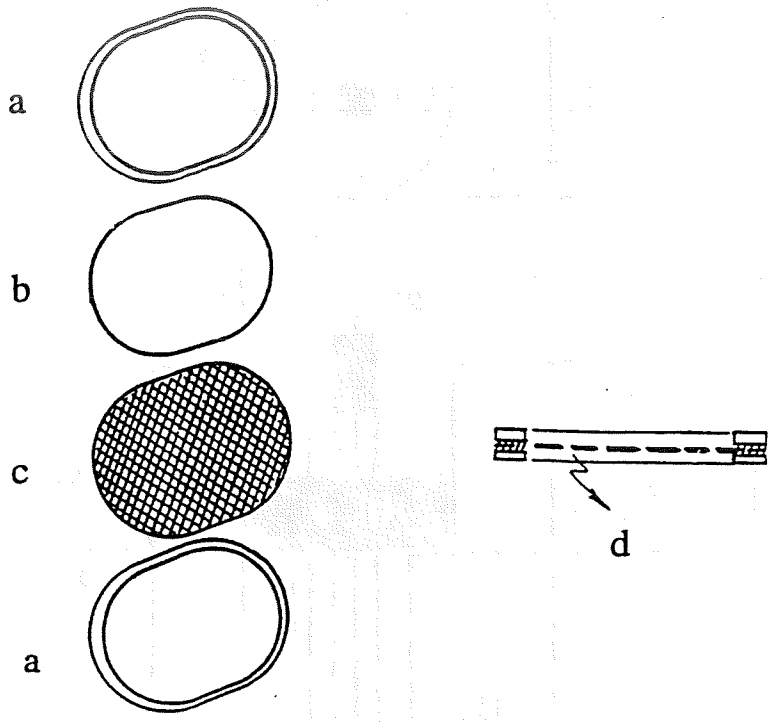


FIGURE 3. Detail of platinum screen anode for regeneration of tris[2,2'-bipyridine]ruthenium(III) reacting centers. a: Platinum rings; b: cellulose-based filter material; c: platinum screen; d: Nafion-immobilized tris[2,2' bipyridine]ruthenium(II) layer.

sensing device. The semi-solid layer of Nafion-complex in contact with the minigrid platinum working electrode is in direct contact with the cell compartment in which the reference and auxiliary electrodes are located. This compartment can be filled with aqueous supporting electrolyte solution which is in contact with the Nafion-complex layer via the central orifice machined at the bottom of the upper body of the unit. Gravitational flow provides a continuous path for compensating the water loss produced by air-drying. This strategy facilitates electrical as well as ionic communication, both required for the operation of the Ru(II)/Ru(III)/Ru(II) cycle:

hydrazine + Ru(III)

$\{\ \}\ e^-$

Ru(II)

This arrangement also provides convenient segregation of the working electrode from the auxiliary one and of the electrochemical and optical parts of the sensor. It requires, however, a compromise between ionic mobility and percolation of the liquid to keep the screen wet. For this reason the intercalation of cellulose-based filter pads (see Figure 3) was implemented.

II. MATERIALS AND METHODS

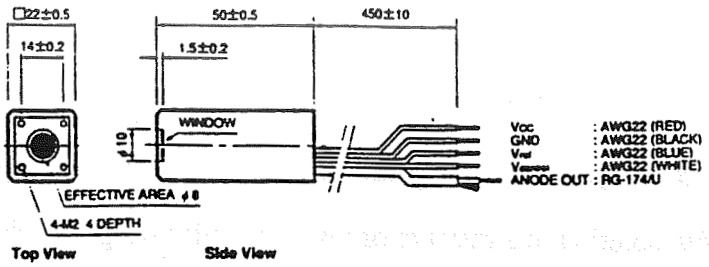
2.1. INSTRUMENTATION

The prototype sensor was constructed of Teflon and black Delrin, and essential details are shown in Figure 2. The platinum rings and platinum screen were cut and forced molded from the reinforcement rim and mesh body of a discarded platinum electrode similar to the type used in electrogravimetric determinations. The reference electrode was a R-5 silver/silver chloride minielectrode with Vycor tip and 3.0 M NaCl filling solution (Bioanalytical Systems, West Lafayette, IN). The Teflon tubing (1 mm i.d.) used for sample transport into and out of the sensor as well as the tube end fittings were of the type used in liquid chromatography and were obtained from Upchurch (Oak Harbor, WA). The optical window was a 0.10 mm thick borosilicate glass circular microscope cover.

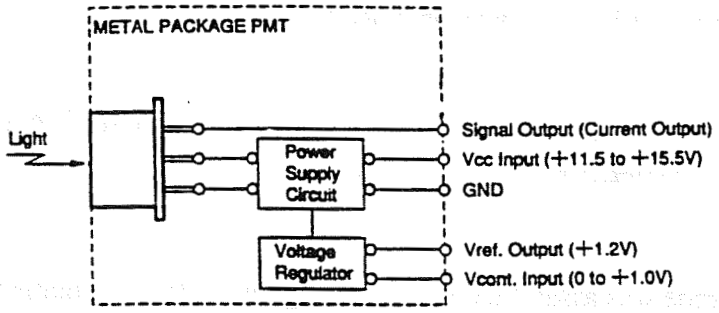
Cyclic voltammetric information was collected with the aid of a CV-1B Voltammograph connected to a RXY-MF8050 recorder. Amperometric measurements were performed with the help of a LC-4B potentiostat/amplifier, all obtained from Bioanalytical Systems (West Lafayette, IN).

A photosensor module type H5783-01 from Hamamatsu Photonics K K. (Bridgewater, NJ) with broad spectral response (300 to 820 nm range) was attached to the lower part of the cell (just in front of the glass window shown in Figure 2) for light detection. This type of detector exhibits high sensitivity, wide dynamic range, and fast response. Dimensional outlines, module functional diagram, and wiring for sensitivity adjustment are shown in Figure 4. The +11.5 to 15.5 V power supply for sensitivity adjustment was designed by

Dimensional Outline (Unit: mm)



Module Functional Diagram



Wiring For Sensitivity Adjustment

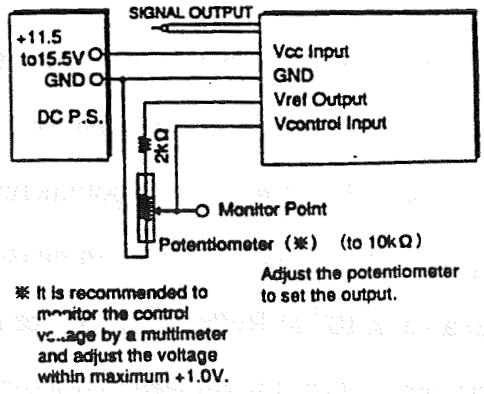


FIGURE 4. Dimensional outline, module functional diagram, and wiring for sensitivity adjustment (resistance programming) for the Hamamatsu H5783-01 photosensor module.

Carl Mattson and built by David Counts of the Toxic Vapor Detection/Contamination Control Laboratory.

2.2. CHEMICALS AND SOLUTIONS

All chemicals were of analytical reagent grade and used as received, except as noted. Nafion as a 5% (w/w) solution in a mixture of low molecular weight aliphatic alcohols containing 10% water or in a Teflon reinforced membrane (0.007 inch thick) was obtained from Aldrich (Milwaukee, WI). Perchloric acid, sodium perchlorate, 2,2'-bipyridine and the 2,2'-bipyridine rutheneous dichloride were from GFS Chemicals (Columbus, OH). Ruthenium(III) chloride trihydrate was obtained from Aldrich.

2.3. PREPARATION OF THE NAFION-IMMOBILIZED $Ru(bpy)_3^{2+}$ POLYMERIC FILM ON TOP OF THE PATINUM SCREEN

The working electrode was assembled by cementing one of the two platinum rings to the copper contact (C in Figure 2) with a fine suspension of silver metal in butyl acetate (GC Electronics,Rockford, IL, distributed by Allied Electronics, Fort Worth, TX) and allowing the organic solvent to air-evaporate. This silver print provides good electrical conductivity, keeps the ring in place, and protects the copper contact. A Nafion solution was deposited onto the screen with the aid of a small brush (of the type used for painting in toy-model construction) until it appeared covered by a fine layer of polymeric film. After air-drying, helped with the use of a heat gun, the film was swollen by immersion of the screen for a minimum of 2 to 4 hours into a 1.0×10^{-3} M $Ru(bpy)_3^{2+}$, 2 Cl^- or $Ru(bpy)_3^{3+}$, 3 Cl^- aqueous solution (on occasions this contact was allowed to last overnight). The assembled working electrode (see Figure 3) was then incorporated into the cell. The screen was kept moisten with electrolyte solution by inserting two or three circles of cellulosic cloth filter (19 mm in diameter and 1 mm thick) between the screen and the upper cell body.

III. RESULTS AND DISCUSSION

3.1. CYCLIC VOLTAMMETRIC STUDIES

Cyclic voltammetry is a powerful technique for the characterization of electron exchange processes taking place at electrode surfaces [9]. Figure 5 shows the typical cyclic voltammogram for $Ru(bpy)_3^{2+}$ immobilized on Nafion. For thin layer electrochemistry of immobilized species performed under reversible electrochemical conditions (fast electron exchange) the anodic and cathodic peak potentials coincide; with the conditions used here a peak separation of 0.103 V indicates a quasi-reversible behavior (i.e. a moderate slow rate of electron exchange). As expected, peak heights increased with increasing scan rate from 10 to 200 $mV.s^{-1}$, and peak separation also increased with increasing scan rate. A plot of cathodic peak current as a function of the square root of the scan rate follows a linear trend with regression coefficients in the 0.975 to 0.990 range, as it should be expected. The voltammogram of Figure 5 was obtained by casting a film of Nafion-immobilized $Ru(bpy)_3^{2+}$ on a small platinum disk electrode. Voltammograms at scan rates higher than150 $mV.s^{-1}$ developed cathodic and anodic currents out of scale with the instrument used here.

Films deposited on the platinum screen, and assembled as the working electrode in the prototype sensor, exhibit, however, voltammograms as the one illustrated in Figure 6. The appearance of a single reduction peak indicates: (a) the electrochemical process tends toward "irrevers.bility", i.e. the electron-exchange mechanism is even slowe: on the screen than on a small platinum electrode directly immersed into a solution of supporting electrolyte (this may be the result of the geometry of the cell, and/or a relatively poor ionic communication between the cell reservoir and the platinum screen), and (b) oxidation does take place during the anodic scan, and $Ru(bpy)_3^{3+}$ forms in the vicinity of the platinum filaments constituting the screen. Direct visual observation of screens kept for 24 h at a

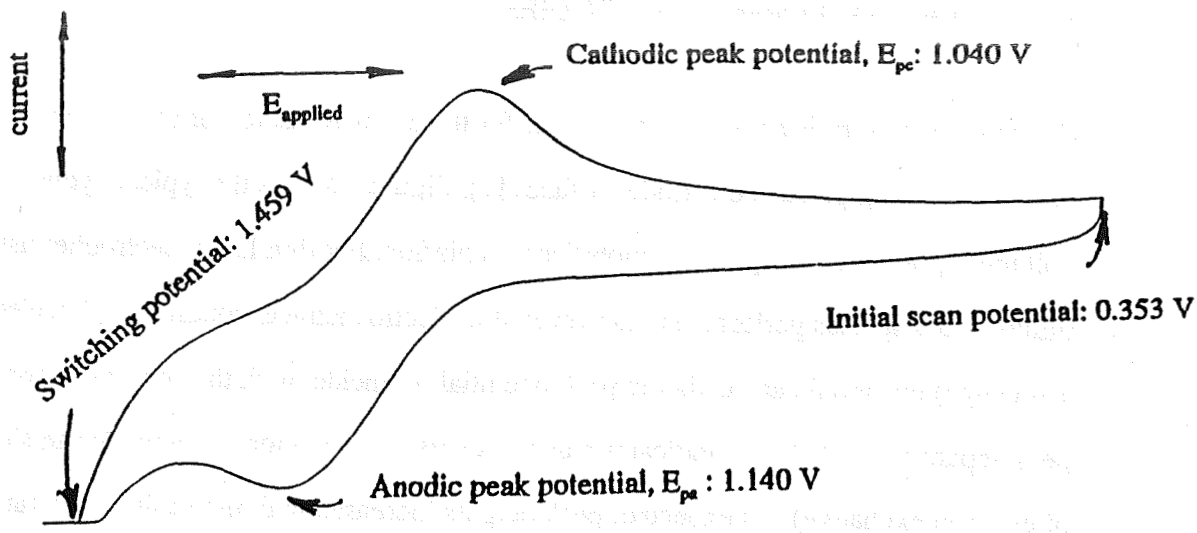


FIGURE 5. Typical cyclic voltammogram of $Ru(bpy)_3^{2+}$ immobilized on Nafion. Platinum disk electrode (2 mm diameter); supporting electrolyte: 0.025 M KH_2PO_4, 0.10 M K_2SO_4, pH = 4.04. Scan rate: 100 $mV.s^{-1}$. Reference electrode Ag/AgCl, 3 M NaCl.

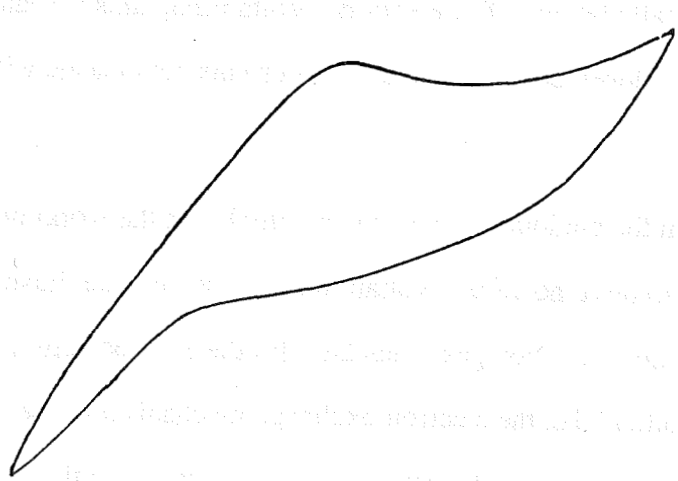


FIGURE 6. Cyclic voltammogram of Nafion-immobilized $Ru(bpy)_3^{2+}$ on the platinum screen of the sensing unit. Supporting electrolyte: 0.025 M KH_2PO_4, 010 M K_2SO_4, pH =4.03. Scan rate: 50 $mV.s^{-1}$. Initial scan potential: 0.148 V. Switching scan potential: 1.54 V. Reference: Ag/AgCl, 3 M NaCl.

potential of +1.25 to +1.42 V vs. the Ag/AgCl, 3 M NaCl electrode, however, did not show evidence of external green $Ru(bpy)_3^{3+}$ formation. This observation and the cyclic voltammogram of Figure 6 suggest that oxidation occurs in the interior part of the polymeric film deposited on the screen. Microscopic observation of a screen oxidized at constant potential did not rule out such a possibility; it rather pointed in such direction in some isolated portions of the screen. This indicates the development of preferential paths for electron transfer within the screen.

In the suspicion that high concentrations of positively charged metal ions present as part of the supporting electrolyte can compete with ruthenium(II/III) for occupancy of the negatively charged sulfonate groups on Nafion, a cyclic voltammetric test was performed with the small platinum electrode coated with a film of Nafion-immobilized $Ru(bpy)_3^{2+}$. The result of this test is shown in Figure 7. Although rather slowly, supporting electrolytes containing rather high ionic strength removed the ruthenium complex from the Nafion film. The rate of leaching, however, does not seem of consequence with electrolytes of moderate ionic strength which are most likely to be used in sensor development.

3.2. STUDY OF IONIC COMMUNICATION BETWEEN THE SCREEN AND THE SUPPORTING ELECTROLYTE RESERVOIR

As indicated earlier, the need for compensation in solvent evaporation and for the sensor unit to keep a wet environment on both sides of the screen requires a compromise between ionic communication and solution percolation. Cyclic voltammetry, again, provided the guiding information. The original design gave very poor ionic mobility impairing the redox process at the platinum screen. As shown in Figure 8A, no clear redox peaks were visible. By increasing the diameter of the hole that provides communication with the cellulosic pads/screen area, or by decreasing the number of cellulosic pads, the reduction peak of oxidized ruthenium became apparent (Figures 8B and 8C). Although the optimun

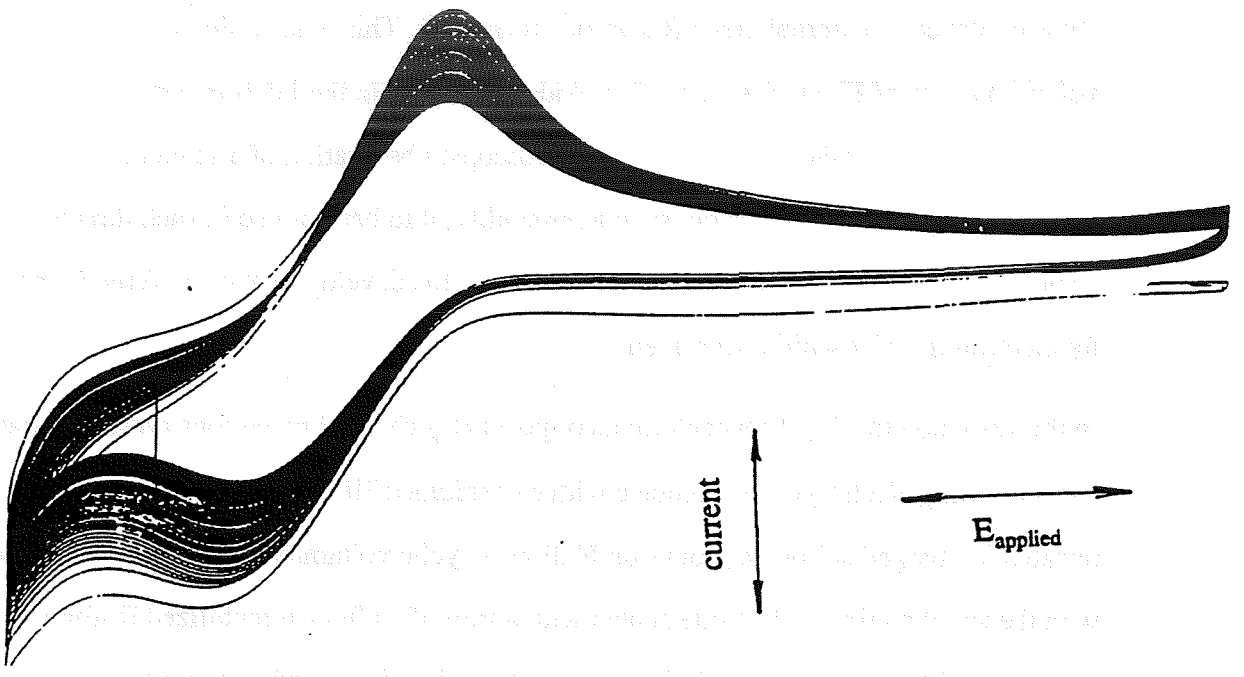


FIGURE 7. Successive cyclic voltammograms of $Ru(bpy)_3^{2+}$ immobilized on Nafion showing the slow removal of complex by a high ionic strength supporting electrolyte (0.10 M KH_2PO_4, 0.50 M K_2SO_4, pH 4.03). Working electrode: platinum disk of 9 mm diameter. Scan rate: 100 $mV.s^{-1}$. Initial scan potential: 0.343 V. Switching potential: 1.435 V. Reference electrode: Ag/AgCl, 3 M NaCl. Leaching is shown by a decrease in current in successive scans. Time between scans: 15 minutes.

compromise was not reached for lack of time, the direction to arrive at such an optimun has been shown.

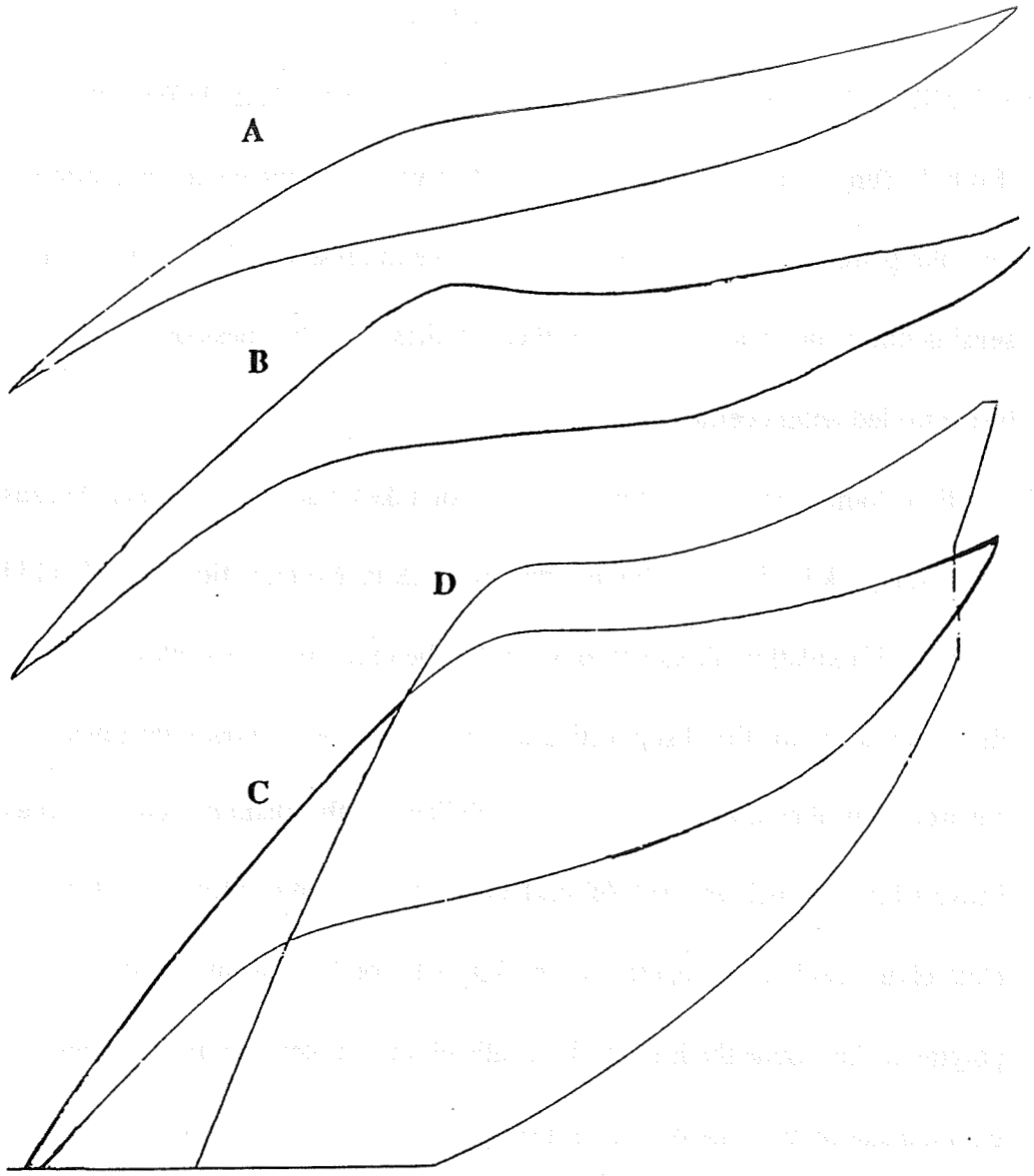


FIGURE 8. Effect of cellulosic filter pads inserted between the bottom of the upper cell body and the platinum screen and the diameter of the perforation at the bottom of the upper cell body which communicates the screen area with the supporting electrolyte reservoir. **A**: 1 mm diameter communicating channel, 4 filter pads; **B**: 1 mm diameter communicating channel, 3 filter pads; **C**: 2 mm diameter communicating channel, 3 filter pads; **D**: 2 mm diameter communicating channel, 2 filter pads. Scan rates: 50 mV. s^{-1}. Initial scan potential: 0.398 V. Switching potential: 1.400 V vs. Ag/AgCl, 3 M NaCl.

IV. CONCLUSIONS

4.1. OBSERVATIONS RESULTING FROM THE PERFORMED WORK

1. Both, $Ru(bpy)_3^{2+}$ and $Ru(bpy)_3^{3+}$ immobilize very well and easily on Nafion deposited on solid platinum surfaces as well as on the platinum screen that is part of the prototype sensing unit which was the focus of these studies. The thickness of polymeric films can be controlled within certain limits.

2. Cyclic voltammetry shows that: (a) oxidation takes place on the screen, because no oxidation peak is observed, but a reduction peak is. No reduction peak should be observed if oxidation did not take place, (b) the redox process is slow (polymer films in direct contact with stirred supporting electrolyte solutions show quasireversible electrochemical behavior, and films immobilized on the platinum screen revealed irreversible electrochemistry), (c) high concentrations of positive ions (from supporting electrolyte solutions) compete for the $-SO_3^-$ sites on Nafion, and remove from the polymeric backbone the immobilized ruthenium complex, and (d) the ionic communication with the present sensor geometry is not fast enough and may be the rate determining step responsible for the observed irreversible electrochemistry (increasing the size of the communicating channel and/or the flow of supporting electrolyte increases the apparent overall rate of electron exchange).

3. The oxidation of all ruthenium(II) centers (when $Ru(bpy)_3^{2+}$ is the immobilized species) has been unsuccessful. No green color was visually observed when keeping

overnight at a constant applied potential of +1.38, +1.42, +1.55, and even +2.00 V vs. a Ag/AgCl. 3 M NaCl electrode as reference. Potentials of +1.55 and +2.00 revealed considerable oxidation around the copper contact (point C in Figure 2) and the electrolysis of water. Typically, currents of a few tenths of a mA were detected during the oxidation process. This was observed even with very thin polymeric films. However, after an oxidation attempt, and in the process of removing the film, the polymeric material close to the screen appeared pale green. The final portions of wahing liquid (a 1:3 solution of 20% nitric acid and acetone) were also of very pale greenish coloration). This and the cyclic voltammetric observations listed above lead to the conclusion that oxidation is limited to the interior part of the polymeric film in contact with the platinum screen. This seems consequential of a very slow charge hopping within the polymeric film.

4. The $Ru(bpy)_3^{3+}$ complex ion can be prepared reacting $RuCl_3$ with a large excess of 2,2'-bipyridine in aqueous solution (final pH between 3 and 4).

5. The idea of amperometric monitoring and electrochemical regeneration of Ru(III) centers is limited to very thin films. Consequently, a sensor based on the strategy studied here would provide very low limits of detection, but restrict the linear response to concentration change. A sensor can be built using $Ru(bpy)_3^{3+}$/Nafion films prepared by chemical means. Low limits of detection and wide dynamic concentration range should be accessible by this route. The films will require, however, periodic chemical restoration of the ruthenium central metal ion to the III oxidation state.

4.2. SUGGESTIONS FOR FURTHER WORK

1. Verification that Nafion-immobilized $Ru(bpy)_3^{3+}$ prepared by direct reaction between Ru^{3+} and 2,2'-bipyridine produces chemiluminbescence upon reaction with hydrazine.
2. Preparation of very thin films for efficient electrochemical regeneration of Ru(III) centers.
3. Optimization of ionic communication/flow of supporting electrolyte solution and pH optimization (compromised pH) for simultaneous operation of both chemiluminescence and electrochemical processes. The compromised pH is needed because the electrochemical process is favored by a pH of or below 4 and the chemiluminescence process operates better at pH values higher than 7.

4.3. ALTERNATIVES OPEN TO CONSIDERATION

1. Search for an inert support other than Nafion for anchoring of the ruthenium complexes. This inert support should provide reasonably fast charge transfer conditions.
2. Explore the development of a hybrid sensing strategy based on a segmented continuous-flow approach of alternate liquid and gas plugs. This approach should permit operation at the optimum pH for chemiluminescence as well as electrochemical restoration of Ru(III) centers.

V. REFERENCES

[1] Eastman Kodak Company, *Material Safety Data Sheet, Accession Number 900575*, Rochester, NY, 1988.

[2] *Documentation of the Threshold Limit Values and Biological Exposure Indices*, 6th. edn., ACGIH, Cincinnati, OH, 1991.

[3] Ingle, Jr., J. D. and Crouch, S. R., *Spectrochemical Analysis*, Prentice-Hall, Englewood Cliffs, NJ, 1988, Chapter 15.

[4] Hercules, D. M. and Lytle, F. E., "Chemiluminescence from Reduction Reactions", *Journal of the American Chemical Society,* Vol. 88, pp. 4745-4746 (1966)

[5] Noffsinger, J. B. and Danielson, N. D., "Generation of Chemiluminescence upon Reaction of Aliphatic Amines with Tris(2,2'-bipyridine)Ruthenium(III)", *Analytical Chemistry*, Vol. 59, pp. 865-868 (1987), and references 2-6 in this publication.

[6] Lee, W.-Y. and Nieman, T. A. "Evaluation of Use of Tris(2,2'-bipyridyl) ruthenium(III) as a Chemiluminescent Reagent for Quantitation in Flowing Streams", *Analytical Chemistry*, Vol. 67, pp. 1789-1796 (1995).

[7] *G. G. Hawley's Condensed Chemical Dictionary*, 11th. edition, revised by N. Irving Sax and R. J. Lewis, Sr., Van-Nostrand-Reinhold, New York, 1987, p. 805.

[8] (a) Rubinstein, I. and Bard, A. J. "Polymer Films on Electrodes. 5. Electrochemistry and Chemiluminescence at Nafion-Coated Electrodes", *Journal of the American Chemical Society*; Vol. 103, pp. 5007-5013 (1981) (b) Downey, T. M. and Nieman, T. A., "Chemiluminescence Detection Using Regenerable Tris(2,2'-bipyridyl)Ruthenium(II) Immobilized in Nafion", *Analytical Chemistry*, Vol. 64, pp. 261-268 (1992).

[9] Bard, A. J. and Faulkner, L. R., *Electrochemical Methods, Fundamentals and Applications*, Wiley, New York, 1980, Chapter 6.

1995 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

514-74
7754
p. 28

SCHLIEREN OPTICS FOR LEAK DETECTION

Dr. Robert E. Peale, Assistant Professor
Mr. Patrick L. Summers, Student
Physics Department
University of Central Florida
Orlando, Florida

Mr. Alranzo B. Ruffin, Student
Physics Department
University of Michigan
Ann Arbor, Michigan

KSC Colleague - Carolyn McCrary
Instrumentation and Hazardous Gas Monitoring

Contract Number NASA-NGT-60002
Supplement 19

August 10, 1995

PAGE 394 INTENTIONALLY BLANK

ABSTRACT

The purpose of this research was to develop an optical method of leak detection. Various modifications of schlieren optics were explored with initial emphasis on leak detection of the plumbing within the orbital maneuvering system of the space shuttle (OMS pod). The schlieren scheme envisioned for OMS pod leak detection was that of a high contrast pattern on flexible reflecting material imaged onto a negative of the same pattern. We find that the OMS pod geometry constrains the characteristic length scale of the pattern to the order of 0.001 inch. Our experiments suggest that optical modulation transfer efficiency will be very low for such patterns, which will limit the sensitivity of the technique.

Optical elements which allow a negative of the scene to be reversibly recorded using light from the scene itself were explored for their potential in adaptive single-ended schlieren systems. Elements studied include photochromic glass, bacteriorhodopsin, and a transmissive liquid crystal display. The dynamics of writing and reading patterns were studied using intensity profiles from recorded images. Schlieren detection of index gradients in air was demonstrated.

SUMMARY

The research described here consisted of a series of simple experiments designed to determine the suitability of a variety of optical schemes for detecting leaks. All schemes are variations of the classic schlieren method, which enhances imaging of index gradients by blocking undeflected rays. This work can be divided into two parts.

The first is an investigation of so-called zebra schlieren with emphasis on leak detection in the orbital maneuvering system of the space shuttle (OMS pod). A high contrast pattern (e.g. zebra stripes on reflecting cloth) is imaged onto a negative of this pattern (e.g. a Ronchi ruling), so that only rays deflected by a disturbance can pass the image plane. We find that the sensitivity of this technique is highest close to the imaging optics and far from the zebra stripes, precisely opposite of what is desired for the OMS pod situation. Simple considerations of geometrical optics explain this observation. An estimate of the maximum deviation expected from, for example, a He leak permits an estimate of the necessary length scale for the zebra pattern. This scale is so fine that loss of contrast in the image is expected from considerations of modulation transfer efficiency and a simple experiment. Hence, it appears that fundamental physical limitations will prevent this method of optical leak detection from finding use in OMS pod processing.

The second stage of the research concentrated on potential schemes for single ended schlieren. Here, the illuminated high contrast pattern is eliminated and a brightly lit outdoor scene is substituted, but the basic idea is the same. The negative of the distant scene must exist in the image plane, and the significant new idea here is to write this negative using reversible media. Media studied were photochromic glass, bacteriorhodopsin, and a transmissive liquid crystal display. Procedures for measuring the modulation of the written pattern and the dynamics of writing and reading the pattern were developed. Schlieren imaging was demonstrated in the laboratory, suggesting that single ended schlieren is nearly ready for field tests. Prototype designs based on each of the reversible media are suggested. A successful single ended schlieren system will have far reaching applications beyond and including the needs at Kennedy Space Center.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

I
INTRODUCTION

Gaseous leak detection is an important problem at Kennedy Space Center, as it is in any large scale industrial activity. The leaks can be classified roughly into two types according to the geometry of the situation. The first type are those which occur in an enclosed and cluttered environment such as the internal plumbing associated with rocket engines. The second are those which occur in large open spaces, such as overland transfer lines, where remote sensing techniques are conceivable and desirable.

In this work, optical methods were explored for the imaging of leaks in both types of situations. The methods are variations of the classic schlieren optics[1]. Experiments described here begin with standard laboratory schlieren and work toward systems with industrial potential. Familiarity with principles and operating conditions at each step were sought and results were documented.

Many variations of schlieren optics have been developed during its long history[1]. By "standard" schlieren we mean all well established and documented variations. None of these is particularly useful for leak detection in industrial settings such as KSC, where the configuration of gas plumbing is determined by considerations other than convenience for schlieren. Specifically, in cramped and cluttered situations such as the internal plumbing of rocket engines the optics of a standard schlieren head cannot be placed around the test region. At the other extreme, for the situation of overland transfer lines, a standard schlieren head would need dimensions on the order of 100 m to simultaneously monitor the entire region of interest. In some situations, setting up of a schlieren head might be prohibited by safety concerns. Hence, a goal is remote schlieren having optics only near the operator.

II
STANDARD SCHLIEREN

Only one of the many classic schlieren variations will be described. Figure 2-1 presents a schematic of the optical set up. Two optical systems are superimposed. The first is the illumination and imaging system for the Ronchi ruling. This system is composed of a diffuse illuminator, a telescope mirror, and a Ronchi ruling (black stripes on glass with a 50% duty cycle). The diffuse illuminator (5 V lamp below 1/16 inch teflon) uniformly illuminates the lower half of the Ronchi ruling as shown. The Ronchi ruling is placed at the mirror's center of curvature (twice the focal length) and symmetrically about the optical axis so that the illuminated half is below the axis. An inverted image of the illuminated part is formed above the axis with unity magnification and is superimposed on the unilluminated part. By moving the Ronchi ruling perpendicular to the optical axis (vertically in Fig. 2-1) the dark parts of the image can be made to line up with the open stripes of the physical ruling. When this occurs, no light will pass the ruling.

To set this system up, look through the upper half of the ruling with your eye located where the camera is in Fig. 1. Then move the ruling together with your head in a plane perpendicular to the optical axis until the illuminated part of the ruling is observed in the mirror. Now move the ruling and your head along the axis until the virtual image fills the mirror, which will then appear uniformly bright. At this point the ruling is nearly at the center of curvature, and stripes should be observed. Further motion along the axis will cause the stripes to increase in size and separation until a single bright or dark stripe fills the mirror.

The schlieren effect is easiest to explain when a dark field exists. Under these conditions, none of the light from the illuminated part of the ruling is getting past the upper part of the ruling. A phase object, such as an index of refraction irregularity in the gas between the mirror and the observer, can cause rays which pass through it to leave the group of rays which form the image of the lower ruling on the upper one. These deviated rays may fall on an open part of the upper ruling and get through. The second optical system operates with these rays.

The second optical system consists of an imaging apparatus, such as the camera shown in Fig. 2-1. The camera images the test region in front of the mirror onto a ccd. Larger index gradients within the test region permit more light to reach the camera so that the intensity distribution in the recorded image maps out the disturbance itself. In practice, imperfect cancellation of undisturbed

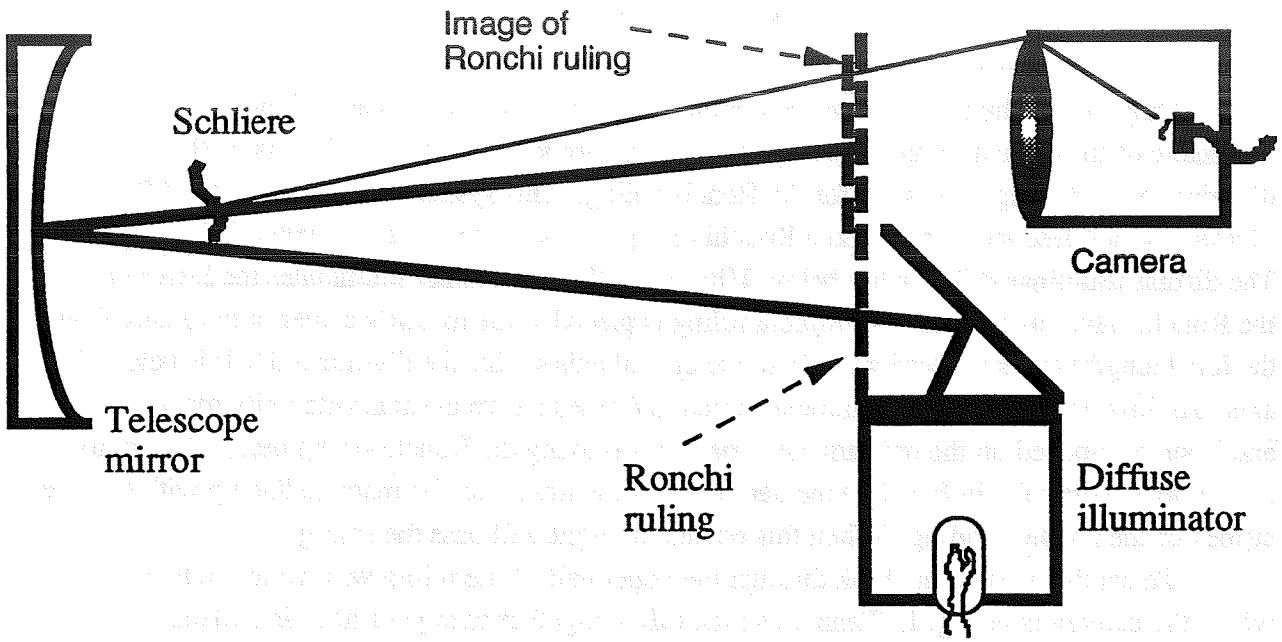


Figure 2-1. Standard schlieren optics.

rays by the first optical system permits an image of the disturbance to be formed with intensity changes of both signs and conveniently allows an image of reference objects to be simultaneously recorded.

Characteristic data collected with the standard schlieren apparatus of Figure 2-1 are presented in Figure 2-2. The data were recorded using an Elmo ccd television camera and the Macintosh based NuVision image processing system by Perceptics. In Fig. 2-2 the disturbance is caused by a squirt from canned air of the type used to dust electronic components. The gas is 1,1,1,2-tetrafluoroethane and has a refractive index significantly different from air. Intensity changes of both signs map out an image of the turbulent flow of gas from the nozzle, located just above and in front of the telescope mirror. None of the disturbance recorded here is visible to the naked eye. With this system, heat waves from warm body parts and even air conditioning currents are easily observed. This standard schlieren system is unsuited for leak detection in industrial settings like KSC because the telescope mirror is too bulky to use behind plumbing internal to rocket engines and too small to encompass the large areas traversed by overland transfer lines.

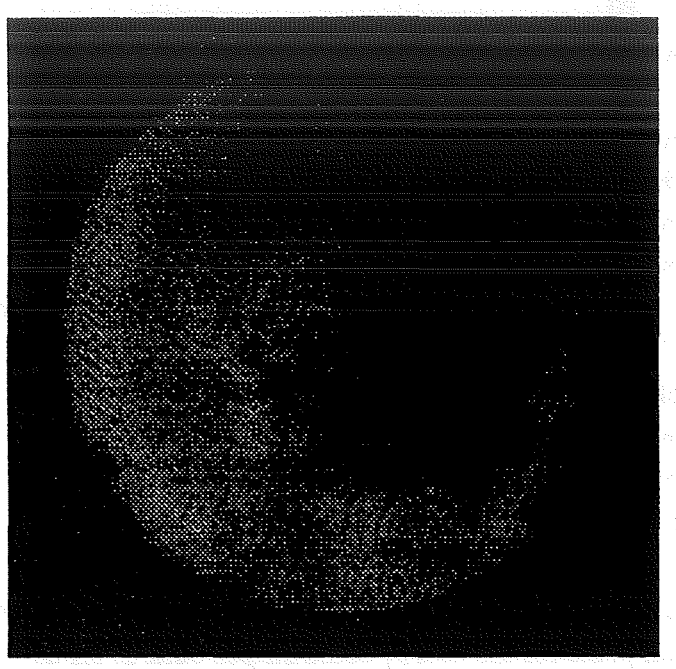


Figure 2-2. Test results of standard schlieren. Canned air is sprayed from the top and turbulence from the expanding jet is observed. Bright bands at the mirror edges are a result of spherical aberration.

III
ZEBRA SCHLIEREN

The following modified schlieren method is intended to eliminate the difficulty caused by the bulkiness of the mirror in situations with limited space, such as the interior plumbing of rocket engines. The name "zebra" comes from the striped pattern which is substituted for the mirror in Figure 2-1. Figure 3-1 presents a schematic of the zebra schlieren apparatus. Stripes made from black construction paper are placed in front of reflecting cloth (3M Corp.). The cloth is illuminated with a projector at close to normal incidence. The imaging optics are also placed close to this axis to take advantage of the strong return of the reflected beam in this direction. A projector is used instead of the diffuse illuminator because the return from the cloth is much less than from the telescope mirror. The zoom lens superimposes an image of the stripes on the Ronchi ruling, and the subsequent theory of operation is similar to that of the standard schlieren. The procedure used to set it up is different however.

Initial alignment is done with the Ronchi ruling removed from the set up. First the focus of the zoom lens is set to infinity to move the focal plane for the nearby zebra stripes back as far as possible. This allows more room for placement of the Ronchi ruling. A simple lens of short focal length is adjusted so that the image formed by the zoom lens is separated from the simple lens by its focal length. Rays leaving this lens are then parallel. The camera is focused at infinity to form an image from these parallel rays. This arrangement allows translation of the camera along the optical axis without losing focus. In particular, the camera can be placed directly behind the simple lens so that the object of interest fills its field of view. Everything is set up correctly when a sharp image of the zebra stripes appears on the TV monitor.

The front of the zoom lens is now blocked and the Ronchi ruling is inserted at the location of the stripe image behind the zoom. A sharp image of the Ronchi ruling should appear on the TV screen. The zoom is unblocked so that the image of the stripes is superimposed on the Ronchi ruling. Moire fringes will be seen in the TV monitor if the image of the stripes has a different scale than the Ronchi ruling. The Ronchi ruling is the correct distance from the zoom when this Moire pattern has its maximum contrast. The Ronchi ruling has the correct orientation when the Moire pattern is parallel with the direction of the stripes. By adjusting the zoom the scale of the image of the stripes can be matched with the scale of the Ronchi ruling. As this condition is approached, the separation of the Moire bands and their width increases until a single dark or light band fills the view. The focus of the zoom may have to be readjusted by monitoring the contrast of the Moire band since changing the zoom causes a change in the position of the stripe image. The Ronchi ruling can now be translated perpendicular to the optical axis to get the appropriate level of extinction. This translation is sensitive and best done with a micrometer stage.

All of the above steps are performed with the simple lens placed so that the camera images the stripes and Ronchi ruling simultaneously. If a disturbance is placed close to the zebra stripes, it will also be in focus with respect to the camera, but will nevertheless be invisible. We find that zebra schlieren is insensitive to disturbances close to the stripes. Placing the disturbance close to the entrance aperture of the zoom gives the best sensitivity, but the simple lens must be moved back so that the image of the disturbance formed by the zoom is at the simple lens's focal point.

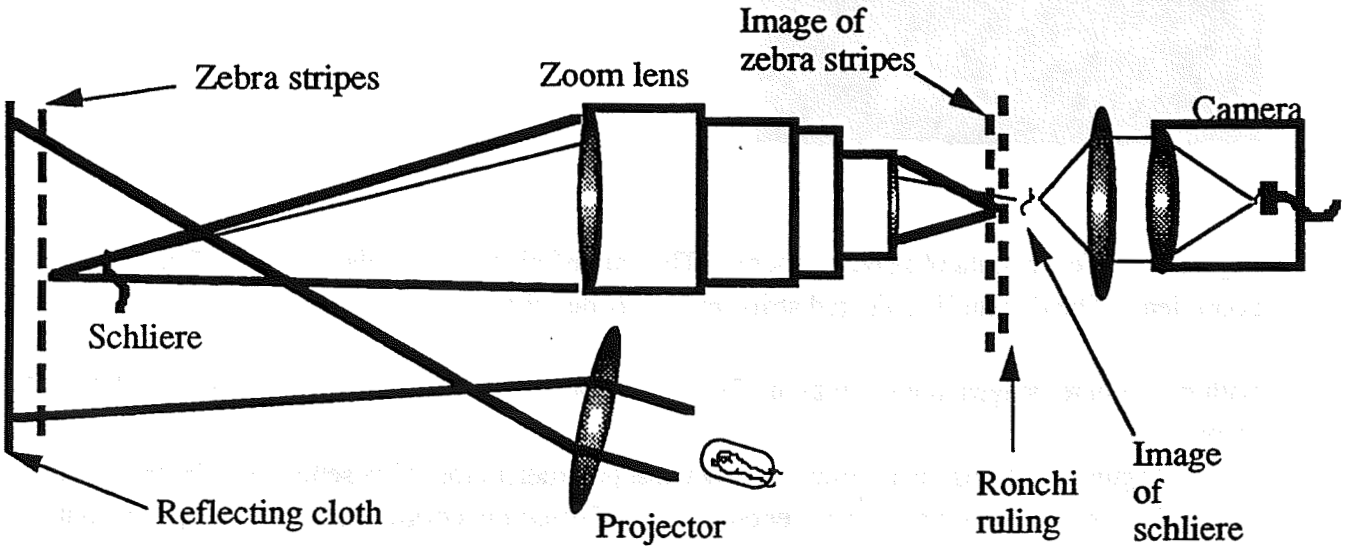


Figure 3-1. Zebra schlieren optics. An illuminated high contrast stripe pattern is precisely imaged onto its negative (Ronchi ruling). Ordinary rays are blocked and deviated rays form an image of the disturbance farther back. This image is viewed with the ccd camera.

Figure 3-2 presents data taken with the zebra schlieren set up. A jet from the inert gas duster is imaged just in front of the zoom. The zebra stripes and Ronchi ruling are out of focus and therefore invisible. The zebra schlieren method studied here is much less sensitive than the standard schlieren, where the turbulent plume of gas can be observed far down stream, as shown in Fig. 1-2. The stream recorded in Fig. 3-2 can also be seen with the naked eye (shadowgraph) though Fig. 3-2 records a significant enhancement. Heat waves from warm body parts were not observed. The actual set up can be significantly improved by fitting the critical optical elements

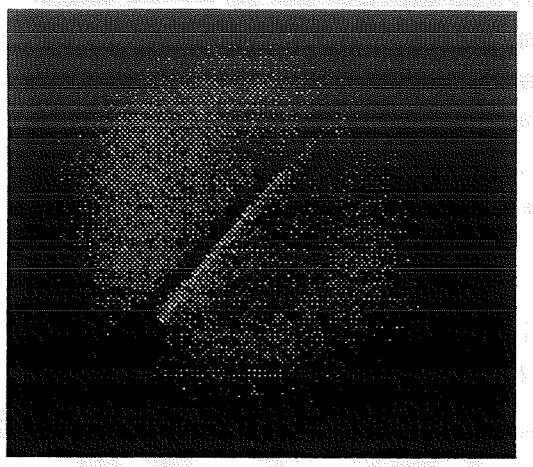


Figure 3-2. Test results of zebra schlieren. The canned air nozzle is placed just in front of the zoom lens. The distant illuminated stripe pattern is out of focus.

with micrometer stages, but the region of sensitivity will still be far from the stripes as will be shown next.

Figure 3-3 gives the pertinent geometrical parameters for zebra schlieren. For the schlieren effect, the disturbance must be large enough that deflected rays from a bright stripe appear to the

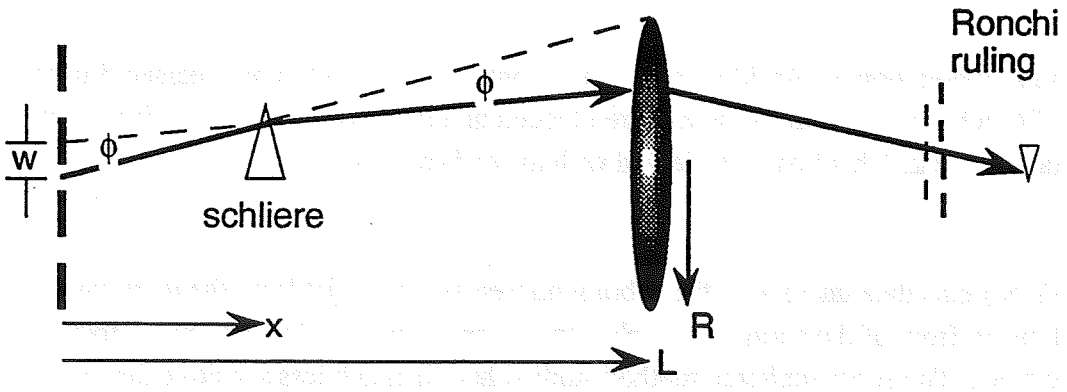
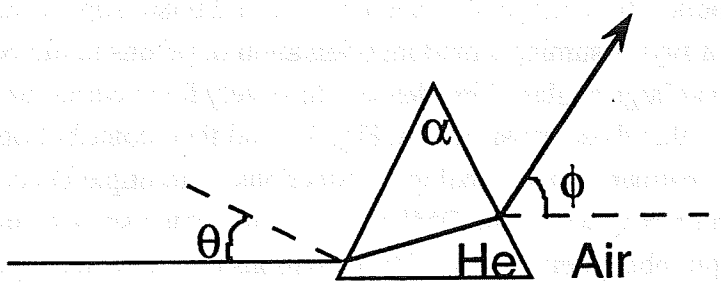


Figure 3-3. Geometrical considerations for zebra schlieren. An image of the stripe pattern is formed in registration with the Ronchi ruling. Rays which pass to form an image of the disturbance appear to subsequent optics to have originated in a dark stripe.

camera to have originated from a neighboring dark stripe. Also the deflection must be sufficiently small that deflected rays are collected by the zoom lens. Both conditions imply a minimum distance x for the disturbance from the stripes for a given deflection ϕ. These are $x \geq w/\phi$ and $x \geq L - R/\phi$.



$$\phi = \theta - \alpha + \sin^{-1}[\sin\alpha\,(n^2 - \sin^2\theta)^{1/2} - \sin\theta\cos\alpha]$$

$$n = n(He) / n(air)$$

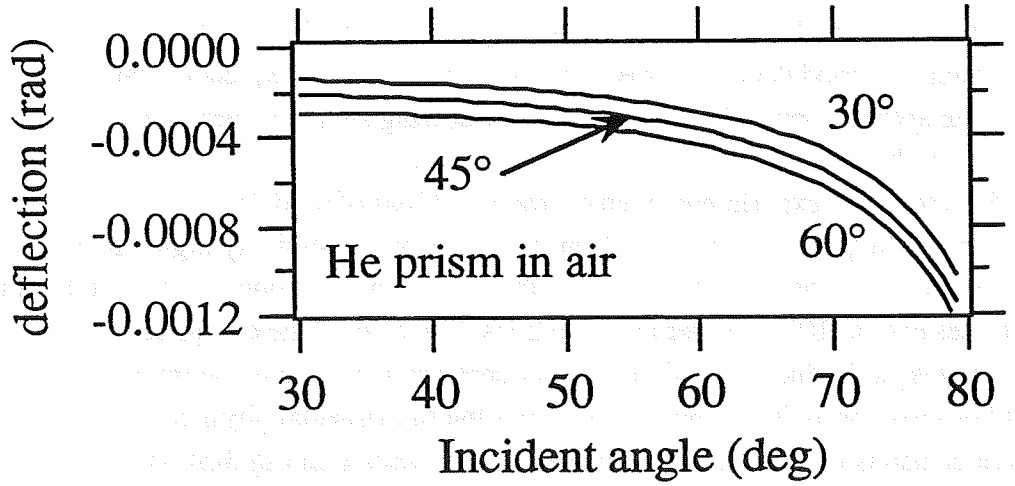


Figure 3-4. Estimate of expected deviation from He leak. A He leak is modeled as a isosceles prism with angle α. The incident angle is θ, and the deflection is ϕ. The formula gives ϕ in terms of α and θ, and the plot presents results for the range of α and θ that the majority of rays are likely to encounter. The maximum deviation is on the order of 1 mrad.

To determine which condition on x is more important for the OMS pod application, where x and L are fixed by external constraints, we must estimate the expected deviation ϕ. Figure 3-4 presents an estimate of the deviation expected from a bubble of He gas in air. The shape of a prism is assumed for simplicity. The formula for ray deviation by a prism is presented and can be found in any elementary optics text[2]. The index of refraction of air is 1.000293. That of He is 1.000036. The sign of the deflection is opposite that of a glass prism in air. The plot in Fig. 3-4 gives the deviation angle in radians for a range of incident angles and prism angles expected to be encountered by the majority of rays assuming a random orientation of prisms in the beam. The maximum deviation is 1 mrad at large angles of incidence. Relatively few rays are expected to have larger angles of incidence that those considered in Fig. 3-4 and their contribution to the effect will therefore be small. This treatment ignores multiple deflections by multiple He bubbles.

The geometry of the plumbing within the OMS pod is such that the stripes can be no farther than about 1 inch behind the plumbing being tested. For deviations of 1 mrad and optics on the order of 1 inch radius, the condition of acceptance of rays by the optics is unimportant for distances $L < 80$ ft. Hence the condition on stripe width prevails under circumstances likely to be encountered. With $x = 1$ inch and $\phi = 1$ mrad, we obtain stripe widths $w \leq 0.001$ inch. This is a very small dimension to work with. Assuming a high contrast pattern of such scale can be produced, the physical limitation is likely to be in imaging this pattern on a Ronchi ruling, which would have to be even smaller. Aside from the difficulty of alignment with the inevitable primary (Seidel) aberrations, there is the loss of contrast caused by diffraction. Rather than consider theoretically the decrease in modulation transfer function at high spatial frequencies caused by diffraction from finite aperture optics[2], we present the following simple experimental demonstration of the problem.

Figure 3-5 presents the experimental demonstration of loss of modulation transfer efficiency at high spatial frequency. Images of Ronchi rulings of successively higher line density are projected on a screen with magnification 3.5. As the data show, the contrast between light and dark fringes decreases until at 100 lines per inch, the lines disappear. (The complete disappearance at highest line density is partly due to additional loss of contrast in recording the image. Some modulation is still observed visually.) This demonstrates the fundamental physical difficulty in imaging fine patterns, and is in addition to aberrations (also observed in Fig 3-5) which cause loss of focus at even small distances from the optical axis. Hence, use of zebra schlieren for leak detection in confined geometries such as the internal plumbing of rocket engines appears unpromising.

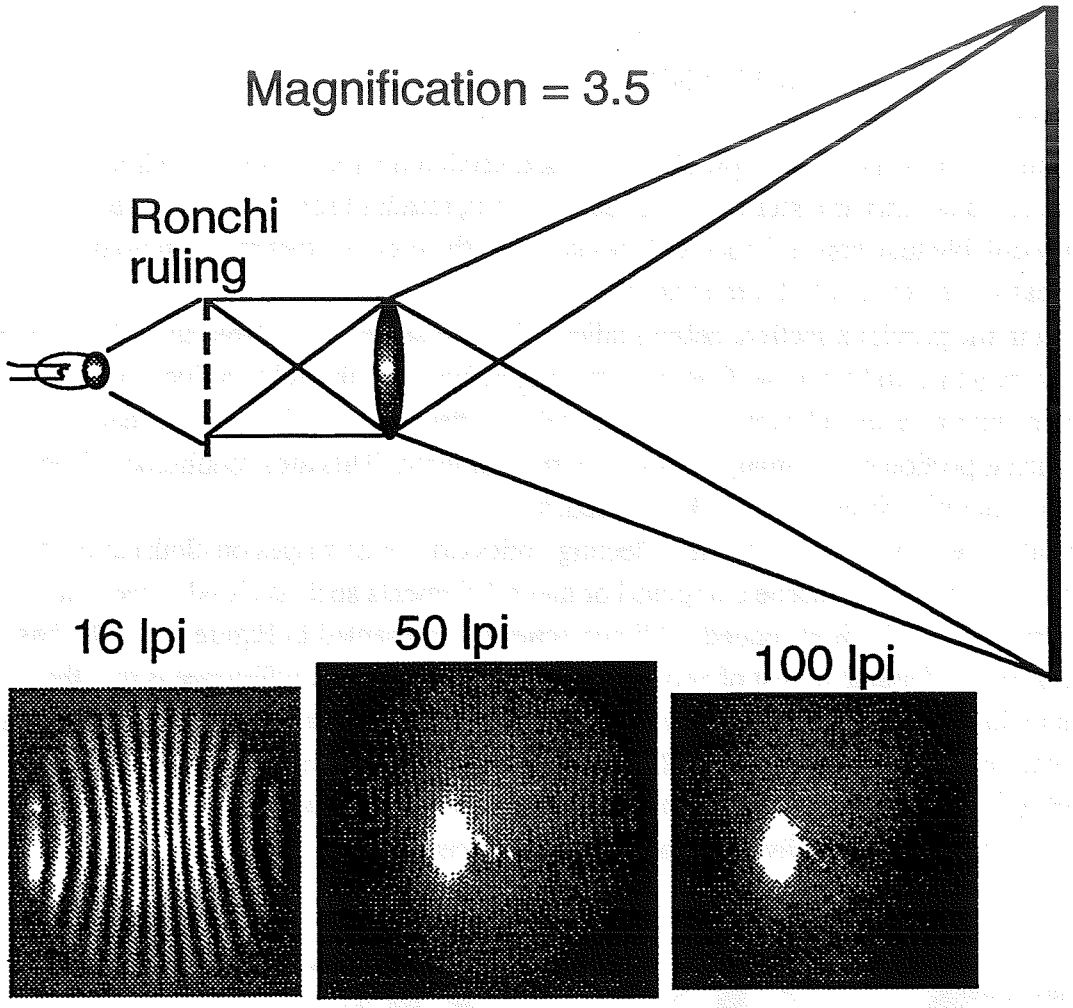


Figure 3-5. Projected Ronchi rulings. This plot demonstrates the well known loss of modulation transfer efficiency with increased spatial frequency. The data are digital photos of the projection screen for Ronchi rulings of different line densities.

IV
SINGLE-ENDED SCHLIEREN

We turn our attention now to the problem of leak detection over large areas, such as encountered with cross country transfer lines. As already said, standard schlieren is limited by the smallness of the available telescope mirrors to dimensions on the order of meters. Simultaneous monitoring of areas on the order of 100 m is desired.

As shown in the previous section, zebra schlieren is most sensitive far from the zebra stripe pattern. If the test region is to be remote from the imaging optics, then the zebra stripe pattern needs to be even further remote. The total area of the zebra pattern needs to be large so that it subtends a significant portion of the imaging system's field of view. This area requirement leads to significant practical difficulties. A new idea is needed.

A potential remedy is to eliminate the reflecting optics (mirror or stripes on cloth) altogether and to use instead a high contrast scene composed of natural elements such as clouds, trees, distant industrial structures, etc. This single ended schlieren scheme is presented in Figure 4-1. The basic principle of operation is identical to that of zebra schlieren. The significant difference is that the Ronchi ruling is replaced by a negative of the distant scene. Photographic negatives are impractical because of the required chemical processing. The ideal situation would be a material which could record the negative while in place within the schlieren system, thus guaranteeing perfect registration, and which could be easily erased to record new scenes.

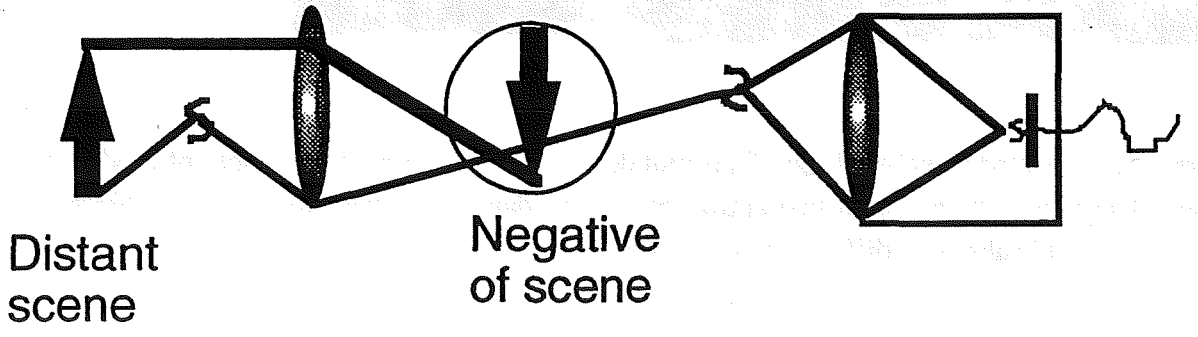


Figure 4-1. Single ended schlieren. The zebra stripes and Ronchi ruling of zebra schlieren have been replaced by a distant outdoor scene and its negative, respectively. In adaptive schlieren the negative is written on an erasable medium.

Three types of material were studied which have the ability to reversibly record the negative of a scene. The first is photochromic glass (Corning photogray) in which a photophysical transformation in microcrystals imbedded in a glass matrix is induced with UV light, causing the material to absorb visible light. The second is bacteriorhodopsin, in which exposure to light of one wavelength causes a reversible increase in absorption at a significantly different wavelength. The third is an electronic method in which a scene recorded with a CCD camera is sent to a transmissive liquid crystal display.

4.1 Photochromic glass

Figure 4-2 presents test results for the photochromic glass using a setup essentially identical to that in Fig. 4-1. The optical system looks into the muzzle of a UV curing gun. One picture is taken immediately after initial UV exposure. The other is after 10 minutes exposure where an obvious darkening is observed. These pictures where collected digitally with the NuVision system, which allows intensity profiles to be determined. The data plotted on the lower left of Fig. 4-2 are such profiles measured across the same part of each image. The contrast plotted on the right is the difference between the two profiles normalized by their sum. With this definition, the range of possible contrast values is 0 to 1. For the photochromic material a value of 0.15 is found. Even with contrast so much less than one, we have observed schlieren with this photochromic glass.

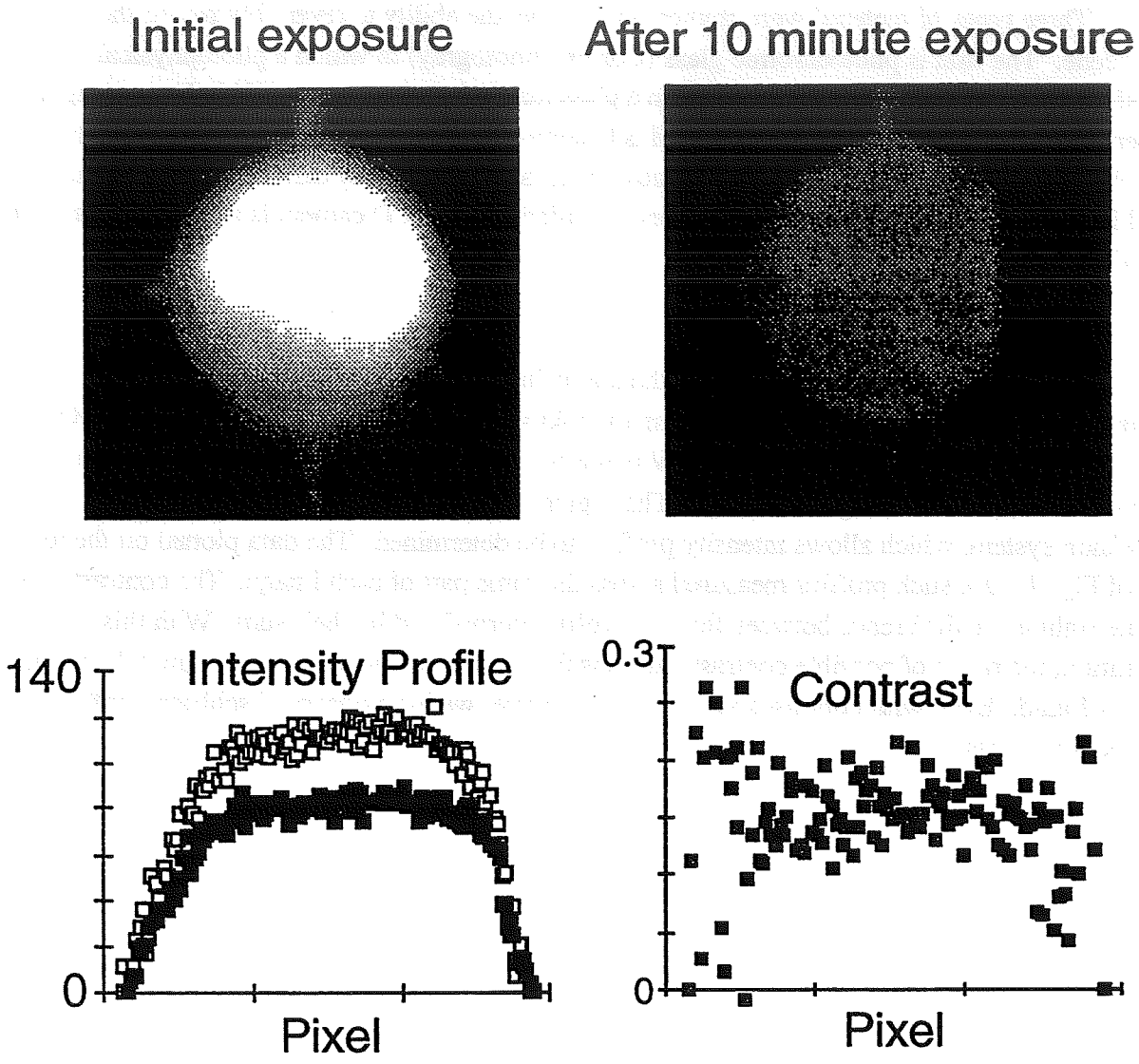


Figure 4-2. Contrast measurements with photochromic glass. The raw data are in the form of images of a UV curing lamp. Pictures are taken upon initial exposure and after 10 minutes. A darkening is evident. Intensity profiles are determined across the image and the contrast is calculated from them as their difference normalized by their sum.

4.2 Bacteriorhodopsin

Figure 4-3 presents a schematic of some of the spectral changes that can be induced in bacteriorhodopsin using monochromatic light[3]. In its ground "bR" state bacteriorhodopsin has an absorption band in the yellow which can be bleached by exposure to yellow light.

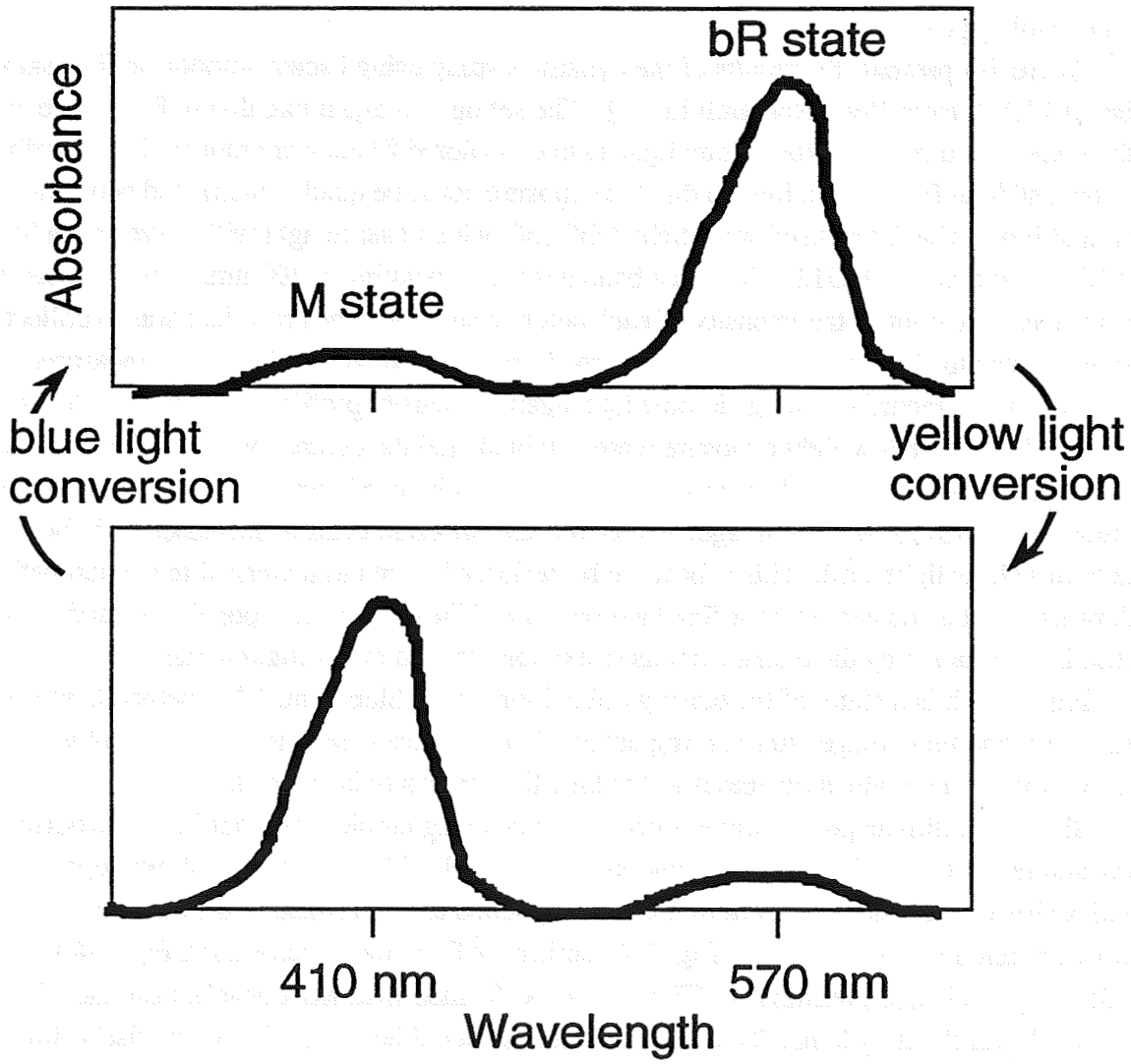


Figure 4-3. Schematic representation of bacteriorhodopsin photocycle.

The bleaching of the yellow band is accompanied by the appearance of an absorption band in the blue spectral region associated with the excited "M" state of the molecule. This phenomenon is the mechanism behind schlieren with bacteriorhodopsin: Exposure to an image in yellow light gives a negative of that image in blue light. One complication is that blue light induces a relaxation back to the bR state. The rate of this reverse reaction is controlled by the intensity of the blue light used. Part of our initial study of bacteriorhodopsin was to uncover the dynamics of the yellow writing and blue reading processes.

Figure 4-4 presents the results of the dynamics study using bacteriorhodopsin film sample number 107190A from Bend Research Inc.[4]. The set up was again like that in Fig. 4-1, except that the scene was a bright diffuse white light source. Colored filters were mounted in a cardboard holder on a slide in front of the film so that the exposure could be quickly alternated between yellow and blue. The filters used were Schott OG550, which passes light with wavelength longer than 550 nm, and Schott BG12, which is a band pass filter peaking at 400 nm. Neutral density filters were used to control the intensity of each color separately. The procedure was to collect an image using blue light, then expose the film to the same image in yellow light for a measured time, and finally quickly record the image in blue light again. Intensity profiles of the blue pictures before and after the yellow light exposure were obtained, and the contrast was calculated as before.

Figure 4-4a shows that the rate at which contrast is induced increases with the intensity of the yellow light used to burn the image. Moreover, the saturated contrast increases with the total intensity of yellow light used. This is because bacteriorhodopsin has a thermal relaxation pathway which proceeds at a constant rate for fixed temperature. The excited state population, and hence contrast, is determined by the balance in this relaxation rate and the excitation rate.

Figure 4-4b is a study of the erasing effect induced by blue light. The method here was to record the second blue image after varying delays. The contrast is seen to decay with blue-light exposure time at a rate which decreases as the blue light intensity is reduced.

Since the ultimate goal is single sided schlieren using ambient outdoor light, we performed writing and reading experiments using outdoor light as well. The door to the lab was opened and the optics aimed at the side of a building, which happened to be in shade. This light was sufficient to both write and read, as the data in Fig. 4-4 confirm. All the measurements in Fig. 4-4 are naturally very crude and preliminary. Timing was performed with wrist watches and the filters were moved in and out by hand. These measurements should be repeated with precisely timed electronic shutters or more elaborate time resolved spectroscopy. Nonetheless, our results describe the basic photodynamics of bacteriorhodopsin.

The maximum contrast obtained in our experiments with bacteriorhodopsin was 0.2. A standard schlieren apparatus similar to Fig. 2-1 was set up and schlieren was observed, as demonstrated in Fig. 4-5. Fig. 4-5a is a blue picture before yellow exposure. A hot soldering iron

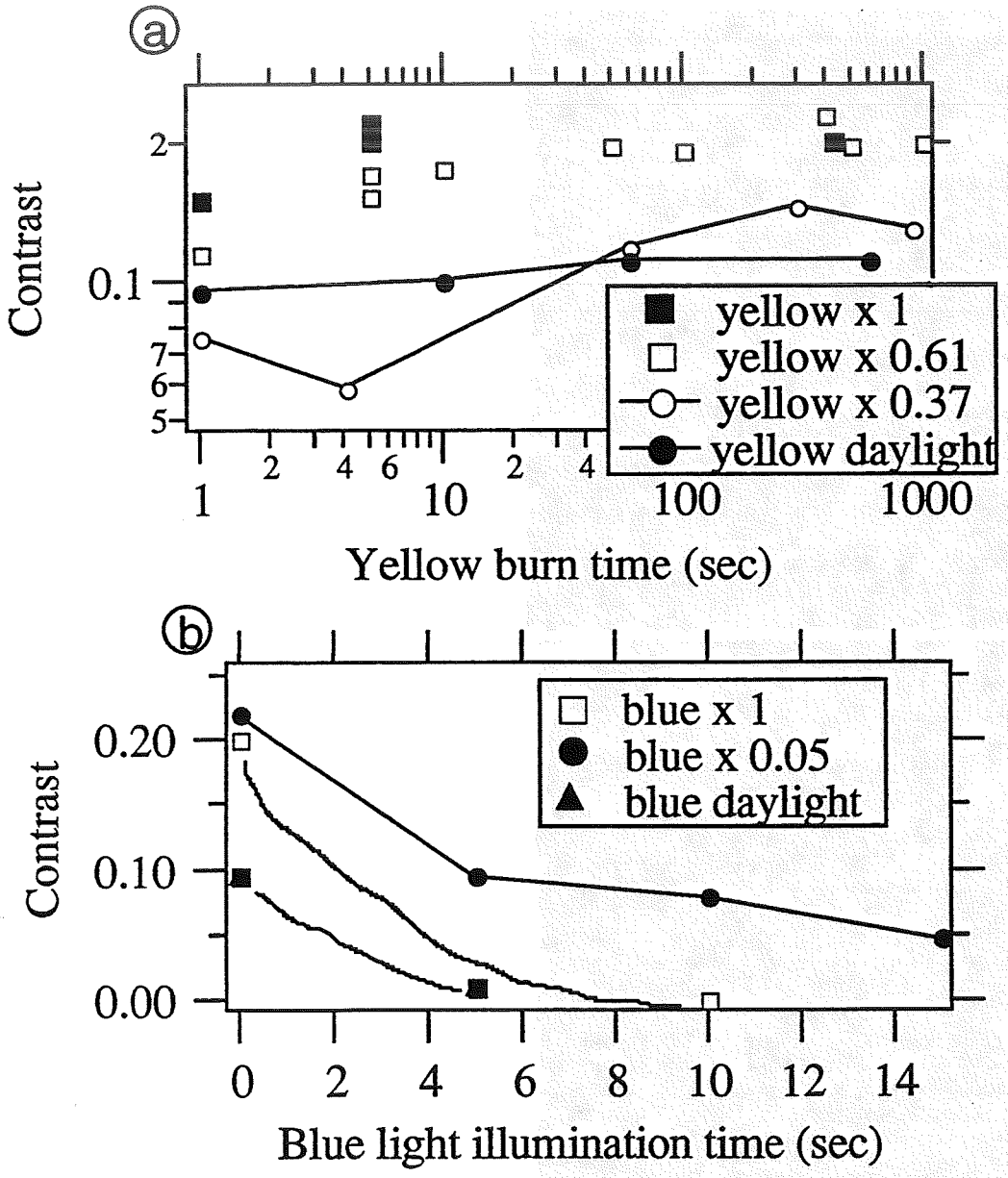


Figure 4-4. Bacteriorhodopsin write and read dynamics. Contrast is defined as the difference in intensity profiles normalized by the sum for images of a lamp taken in blue light before and after yellow light illumination. a) Maximum contrast is plotted against yellow light illumination time for different yellow light intensities. b) Contrast is plotted vs illumination time of the blue read light for different blue light intensities.

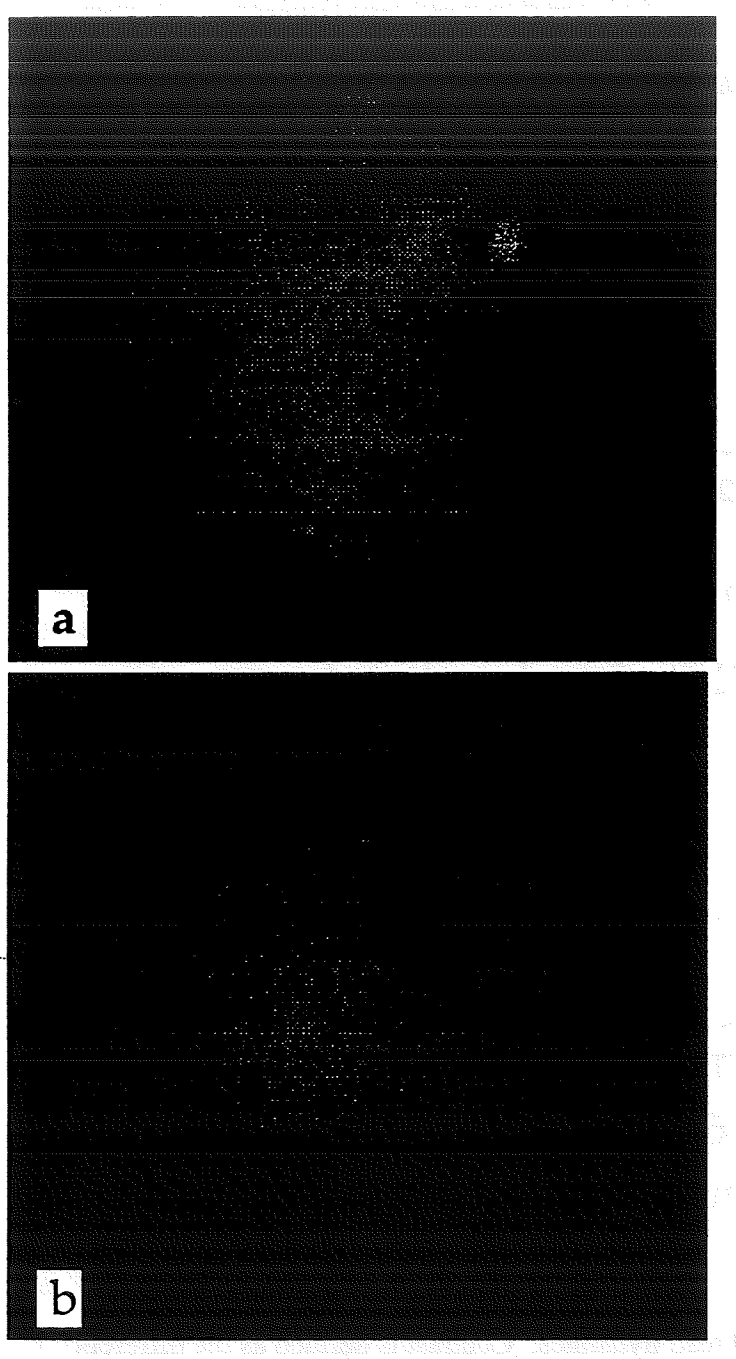


Figure 4-5. Bacteriorhodopsin schlieren test results. A soldering iron is in front of the telescope mirror at the bottom. a) Blue light image taken before yellow light illumination. b) Blue light image taken after yellow light illumination. Index disturbances are observed in the upper right.

is at the bottom edge of the telescope mirror. A 50 line per inch Ronchi ruling is imaged on the bacteriorhodopsin film by the mirror, but this image is out of focus with respect to the camera and is therefore not observed. Some scratches on the mirror surface are observed but otherwise the field is uniform and featureless. Figure 4-5b is a blue light picture taken after the Ronchi ruling has been written into the bacteriorhodopsin film with yellow light. An overall darkening is observed since now the bright stripes of blue are falling precisely on regions of the bacteriorhodopsin which have been made to absorb blue. In addition, the image appears mottled in the upper right quadrant as a result of index gradients induced by heat waves from the soldering iron.

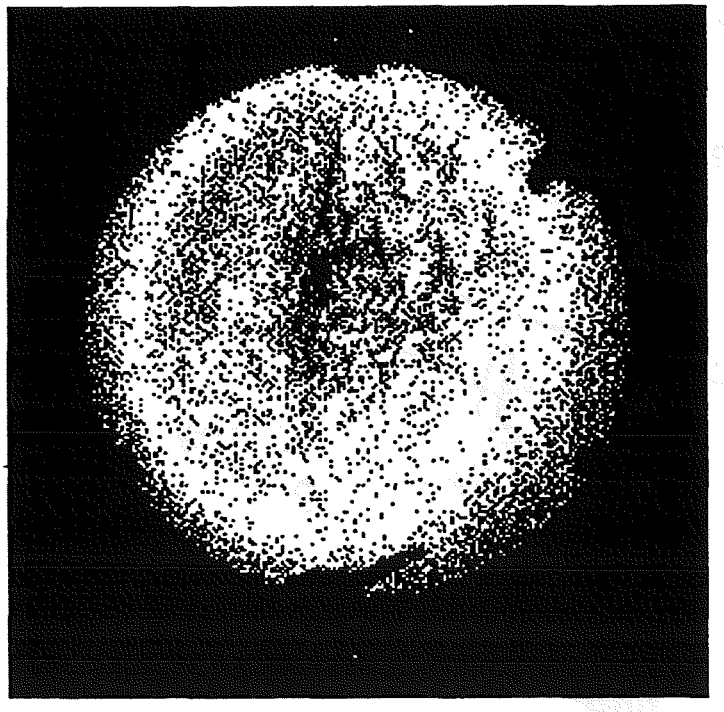


Fig. 4-6. Bacteriorhodopsin schlieren with image processing. Fig. 4-5a has been subtracted from Fig. 4.5b and a binary filter applied to the resulting histogram.

Figure 4-6 shows the same results as Fig 4-5 but with image processing applied. This picture has been included because it is expected to reproduce better than Fig. 4-5. In actuality, Fig. 4-5

contains more information, but this can only be appreciated fully when printed with photographic quality (on Tektronix Phaser IIsdx for example).

To verify that Figs. 4-5 and 4-6 represent a schlieren effect, a control was measured with the Rochi ruling removed. The same exposure steps were performed, except now the light on the film was uniform. The f-stop on the camera, which is behind the film, was stopped down by one to compensate for the two-fold intensity increase caused by removal of the 50% duty cycle Ronchi ruling. Hence, the pictures collected during the control appear to have the same level of brightness as those in Fig. 4-5. (The intensity of light falling on any part of the bacteriorhodopsin film is the same as that which fell in a bright region during the real experiment, but the average intensity entering the camera is doubled). The result of the control was that no effect such as seen in Figs. 4-5b and 4-6 was observed. Hence the Ronchi ruling is essential and Figs. 4-5 and 4-6 are a record of schlieren and not merely shadowgraph.

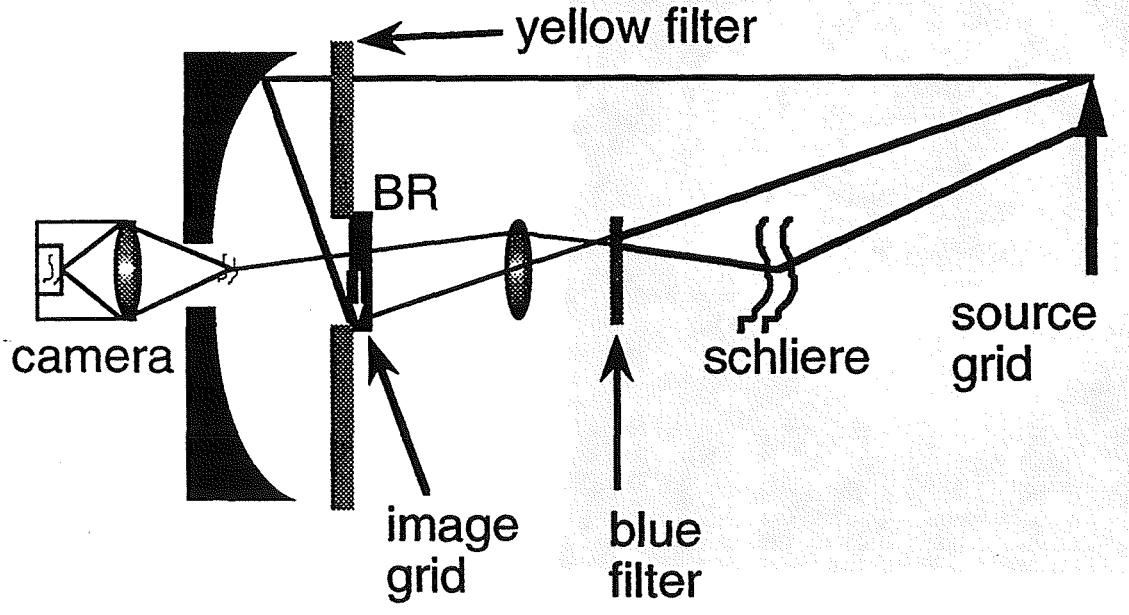


Figure 4-7. Scheme for steady state bacteriorhodopsin schlieren. A negative of the distant scene is continuously written in yellow light from the back side of the bacteriorhodopsin film. The scene is continuously observed in blue light through the front surface. The lens and mirror systems must have the same magnification, so it is convenient to use a zoom lens. Index changes of interest must be rapid compared with the characteristic write time of the bacteriorhodopsin film.

The transient nature of the bacteriorhodopsin write and read cycle is clearly undesirable in a practical device. Figure 4-7 presents a proposed alternative in which yellow write light and blue read light are always present. One disadvantage is that reading and writing is done with independent optical systems, so that automatic registration is lost. A zoom lens should be used to form the blue image. The relative intensities of yellow and blue light will have to be carefully adjusted with ND filters so that disturbances observed in blue are rapid compared with the change in the image stored in the film with yellow.

4.3 Liquid Crystal Display

Figure 4-8 presents the set up for a schlieren experiment using a transmissive LCD display. An illuminated stripe pattern on reflecting cloth is placed at the center of curvature of the mirror but displaced to the left of the optical axis. Initially the LCD is removed from the set up. The mirror forms an image of the stripe pattern with unity magnification to the right of the actual grid. The camera is focused on this image and displays it on the LCD, which acts as a TV monitor. The

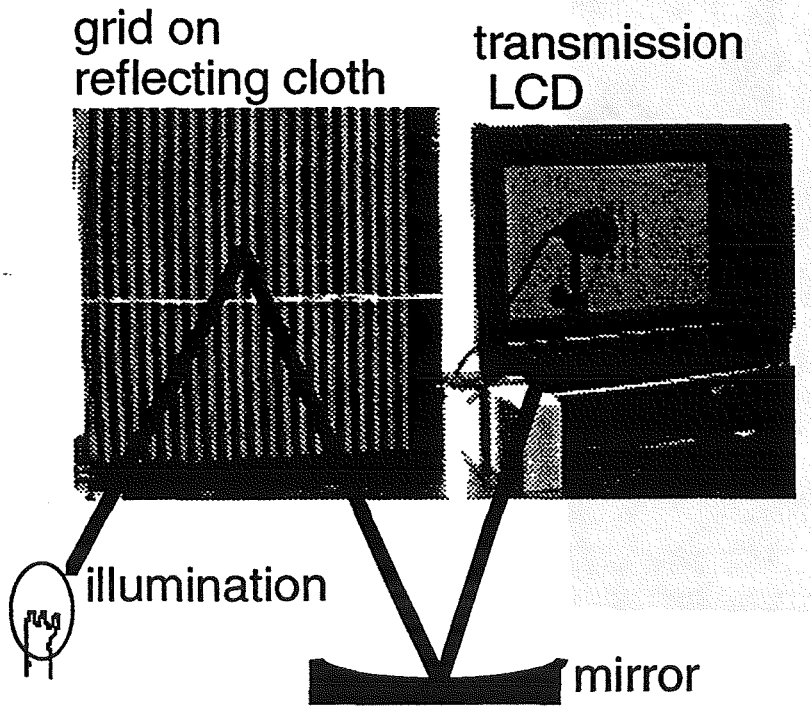


Figure 4-8. Set up for LCD schlieren experiment. The illuminated stripe pattern is imaged with the mirror onto a liquid crystal display, on which a freeze frame image of the pattern is displayed.

camera is moved back from the image plane until the stripes displayed on the LCD are the same width as those on the actual grid and its image. Now the display on the LCD is frozen using a frame grabber, and the LCD is returned to the set up. The camera is moved back close to the image plane so that the LCD picture fills the camera's field of view. The camera now looks through both the LCD and the grid image and is focused on the test region in front of the mirror. The LCD is translated perpendicular to the optical axis to put its displayed stripe pattern in registration with the image of the grid.

Figure 4-9 presents results obtained with the LCD schlieren apparatus just described. Heat waves are clearly visible from the soldering iron in front of the telescope mirror. It is important to realize that even with the LCD removed, a strong shadowgraph is observed which is enhanced by the non uniform illumination coming from the grid. However, sensitivity changes caused by LCD translation verify that a schlieren effect is present.

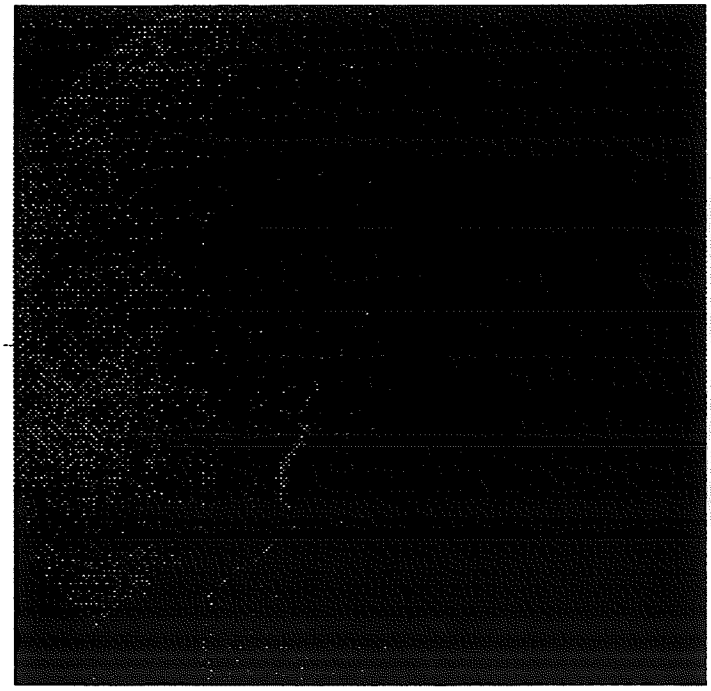


Figure 4-9. Results of LCD schlieren. Heat waves from a soldering iron are observed.

Limitations of the present LCD experiment are the coarseness of the pattern, the lack of sharp edges, limited contrast and resolution of the LCD. However, high-resolution transmissive LCDs do exist for heads-up display applications so that a compact arrangement like that of Fig. 4-1 can be envisioned with higher performance that realized so far.

V
CONCLUSIONS

Variations of schlieren optics have been explored with emphasis on leak detection. For leak detection within confined and cluttered situations (e.g. OMS pod), geometrical and optical considerations indicate a severely limited usefulness for the technique. On the other hand, single ended schlieren using adaptive optical elements such as photochromic glass, bacteriorhodopsin, and electrically controlled liquid crystal displays hold considerable promise as demonstrated by successful preliminary laboratory experiments.

Future work on single ended schlieren will explore other types of photochromic materials, bacteriorhodopsin schemes using continuous illumination and illumination modulated by electrically-switchable color filters [5], and small high-resolution LCDs. In addition, prototypes of actual systems will be developed and tested in field experiments.

REFERENCES

[1] For a collection of historical papers see Selected papers on schlieren optics, edited by Jurgen R. Meyer-Arendt, (SPIE milestone series; v. MS61, The International Society for Optical Engineering, Bellingham, WA, 1992)

[2]. Eugene Hecht and Alfred Zajac, Optics, Addison-Wesley, Reading, MA, 1979.

[3] D. Oesterhelt, C. Bräuchle, and N. Hampp, "Bacteriorhodopsin: a biological material for information processing", Quarterly Reviews of Biophysics **24**, 425-478 (1991).

[4]. Bend Research Inc., 64550 Research Road, Bend, Oregon 97701-8599.

[5] DisplayTech, Inc., 2200 Central Avenue, Boulder, CO 80301

1995 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

S15-16
7755
p. 30

ADVANCED SYSTEMS FOR SHUTTLE LAUNCH PROCESSING

Dr. Rafael A. Perez
Professor
Computer Science and Engineering Department
University of South Florida
Tampa, Florida

KSC Colleagues - Ric Hurt and Rick Dawson
Computer Science

Contract Number NASA-NGT-60002
Supplement 19

July 21, 1995

ACKNOWLEDGEMENT

I would like to express my appreciation to NASA/ASEE for allowing me to spend these ten weeks working on the very interesting and challenging problem of identifying advanced technologies which could be effectively used in the design of a new shuttle launch processing center and to thank Mr. Richard Hurt for selecting me as a member of the group charged with that responsibility. I am very much indebted to Mr. Richard Dawson for orienting me around the Kennedy Space Center, providing me with a place to work, and helping me to focus the information that I gathered on advanced technologies on the unique problems faced by the space program. I would also like to thank other members of the team, Barbara Brown, Ben Bryant, Carl Delaune, Warren Lackie, Chuck Lostroscio, John Moore and James Sudermann for sharing their knowledge with me during all the meetings that we had and for making me feel welcome.

ABSTRACT

Four advanced technologies that could be used in a new shuttle launch processing center are described. The latest methods for high capacity data storage technology, disk arrays and magneto optical disks, are described and their advantages and disadvantages compared. A 3-D protein based optical memory, now being researched, is also described as a possible future technology for data storage. An overview of neural network technology is presented together with several commercial software development options now available for neural network applications. The feasibility of Asynchronous Data Transfer technology as the networking technology to integrate video, voice, and data in a new launch processing center is also considered. Different applications of expert system technology at KSC are enumerated together with a number of commercial expert systems development packages presently available.

SUMMARY

This report describes the following advanced technologies which can play an important role in a new design for the shuttle launch processing center at KSC: high capacity data storage, neural networks, asynchronous data transfer and expert systems.

The latest high capacity data storage technology today consists of disk arrays and magneto optical disks. Each is capable of storing more than the 1.3 Terabytes of data that it is estimated constitute today NASA's requirements for launch related data. Disk arrays are more expensive than magneto optical disks but they provide faster access to data, although it is expected that the access speed for optical disks will decrease within a year or two. Disk arrays make all of the data available online, while a jukebox of magneto optical disks may provide only about 4 gigabytes of online data. It is quite possible that a storage system that combines both of these technologies, i.e. disk arrays for critical data that must be accessed quickly and optical disks for the rest, may provide a good solution for storing data in a new launch processing center.

Neural networks software has been successfully used for a variety of problems. This technology is most effective for problems where the relationships between cause and effect have not been explicitly identified but where there is sufficient historical data from which to learn these relationships. Neural networks can be retrained quickly to learn new relationships and therefore can also be effective for problems that frequently change. A number of software vendors provide reliable and effective development packages for neural network solutions which can execute on Dos, Windows, Mac, or Unix environments. Reasonable hooks from these products to databases are provided so that integration with a data storage system would be feasible.

Asynchronous transfer mode (ATM) is the latest technique developed for communication in computer networks and is expected to become the dominant networking technology in the future. Its main advantages are a transmission speed on the order of gigabits per second and a dedicated switched connection between sender and receiver nodes. These can allow the simultaneous transmission of audio, video, and data within the same network. Presently the main disadvantages of ATM are the lack of standards and the high cost of ATM products. It appears that at the present time most vendors are concentrating on offering ATM as a solution to the problem of limited bandwidth with existing networks. It is rare to find an application which integrates audio, video, and data. Increasing data bandwidth seems to be the bulk of today's demand for ATM products.

Expert systems technology has been successfully used since the 1970's for a wide range of problems. It is considered today by most people to be a proven technology. At KSC there are a significant number of expert systems projects underway or recently completed. A number of reliable software vendors was identified and their brochures provided. As with neural network software, hooks to databases are also available with expert system development software on a variety of operating systems and platforms making integration within an advanced launch processing center very feasible.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

1

INTRODUCTION

A new design for a Shuttle Launch Processing Center requires the identification of advanced systems available today and those that may become available in the near future. The complexity of a shuttle launch processing center is such that many different subsystems need to be carefully integrated into its design. Many different design issues also need to be considered such as: System Architecture, Computers, Networking, Software, Data Storage and Retrieval, Operational procedures, etc.

Identifying advanced technologies that impact all of these design issues is a very large task. This report addresses only the following areas:

- High Capacity Data Storage Technology
- Neural Networks Technology
- Asynchronous Data Transfer Technology
- Expert Systems Technology

The sections below describe the information obtained in each one of the areas.

II

HIGH CAPACITY DATA STORAGE TECHNOLOGY

The approach taken to obtain information about this technology was to first contact reputable commercial vendors of data storage technology and request information on the latest product they sold as well as new products that they planned to introduce in the near future (one or two years ahead). Second, research labs were contacted to find information on products still in the research or prototype stage. The following vendors were contacted. Telephone numbers are given to make it easier to make future contacts with these vendors if needed:

- ANDATACO
  800-3349191 John Aiken
  http://www.andataco.com
  Very good web page. Register first, and the select "tech info" and within that select "raid disk arrays".

- FUJITSU
  800-6264686

- HEWLETT PACKARD
  800-6377740
  http://www.hp.com
  Very good web page. Select "products" then select "computing" then select "information storage".

- HITACHI
  800-4482244 415-2447515 Toni Maglaya

- IBM
  800-4262255
  http://www.ibm.com

- MAXOPTICS
  800-8483092 Less Jones Customer Service
  408-9549700 Akyra Pagoulatos Marketing Manager
  no web page

- SUN
  813-2897228 George Perkins
  407-3800058 Phil Murray
  http://www.sun.com
  Very good web page. Select "product overview" then select "mass storage and peripherals" then select "disks".

- TOSHIBA
  800-9994273 714-4550407

According to these vendors and other technical reports (provided to my NASA colleages), it is generally agreed that the latest technology for mass data storage today consists of disk arrays and magneto optical disks. Important information about each of these is given below.

2.1 DISK ARRAYS

These consist of a group of disk drives that are considered as one logical storage device and which work in cooperation to improve speed and reliability over individual disk operation. Such a technique is referred to as RAID, an acronym for Redundant Arrays of Independent Disks. This technique is ideally suited for environments with large amounts of data needed on line and where data must be protected from disks failures.

RAID improves the speed with which data can be accessed by storing different portions of data blocks on different disks. When the data is needed it can be accessed in parallel from all the different disks with resulting faster access. Reliability can also be improved by this technology by using block interleaved parity, where a parity bit is associated with every block of data. If a disk crashes data can be rebuilt using the rest of the data on the other disks. Most vendors provide "hot spare" capabilities so that disks can be replaced without having to turn the disk array system off. Most vendors also provide software to make it easier to manage storage space allocation.

A technical detailed description of this technology is provided in the "RAID Report". A copy of this report was included in a separate document given to my NASA colleagues together with all the brochures and other information obtained from the different vendors contacted.

Figure 2-1 shows a typical configuration for a RAID and indicates how the configuration could be extended to provide for 1 Terabyte of data storage.

RAID is available from many of the vendors contacted - Andataco, Fujitsu, Hitachi, IBM, and SUN. Typical disk seek times of 10 milliseconds and latency times of 5 milliseconds were reported in the brochures obtained from the different vendors. The RAID storage capacity can be extended considerably. For example a SUN SPARCcenter 2000 connected to 15 Model 200 array cabinets can provide access to 4.86 Terabytes of data storage. The cost of this type of storage is approximately $0.45 per Megabyte, not including the server. For example a 180 Gigabyte system would cost $80,000 today.

2.2 MAGNETO OPTICAL SYSTEMS

These consist of a group of magneto optical read/write disk drives. These drives can be grouped in "jukeboxes" and considered as one logical storage device to provide large amounts of storage space. Most vendors also provide software to make it easier to manage storage space allocation.

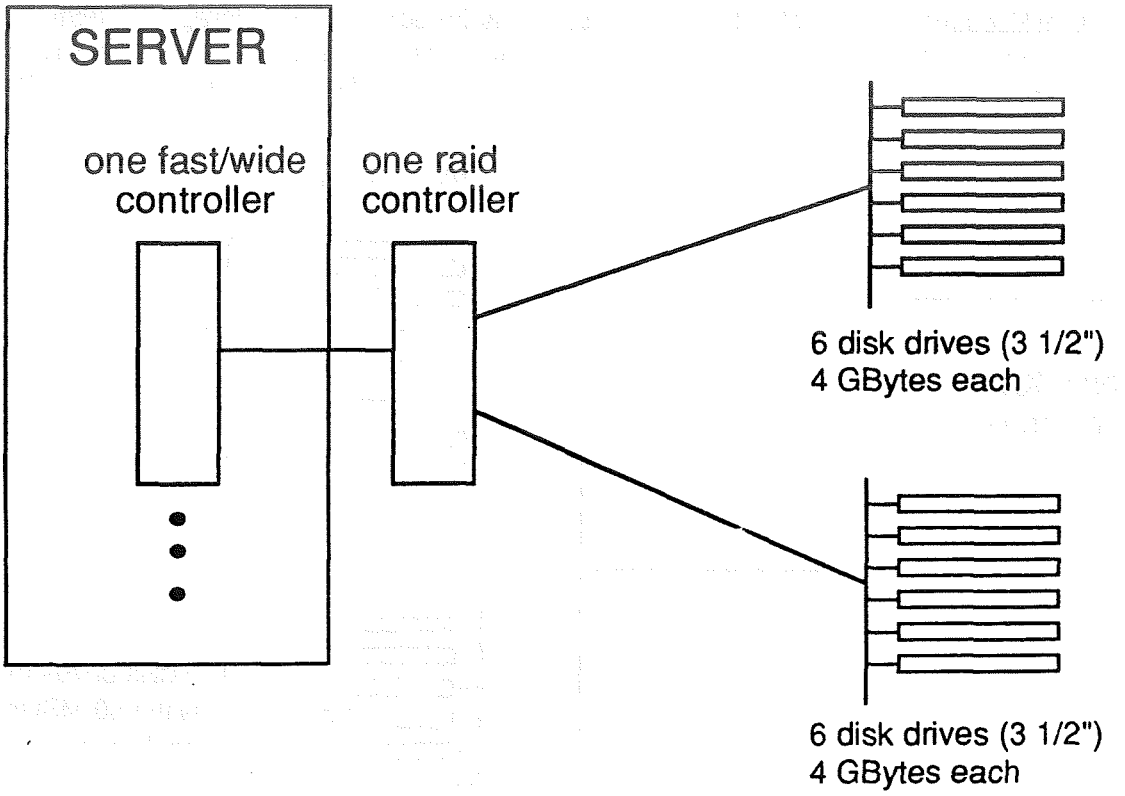


Figure 2-1. Raid Configuration

Magneto optical disks provide low cost on-line access to data and enables data to be stored for very long periods of time. Access speeds, however, are slower that for disk arrays. Presently there are 3 1/2 inch disks capable of storing 230 Megabytes and also 5 1/4 disks capable of storing 1.3 Gigabytes. It is predicted that by the end of 1995, 5 1/4 inch disks will be able to store 2.6 Gigabytes and that by the end of 1996, they will store 5.2 Gigabytes. Notice that since both surfaces on the disk are used to store information, only half of the storage per disk is available online. For example, today, a jukebox with 6 drives would provide approximately 4 Gigabytes of storage online as shown in Figure 2-2. Such a jukebox is capable today of handling 138 optical disks for a total of 180 Gigabytes of storage at a cost of approximately $50,000.

Magneto optical disk storage systems are available from several of the vendors contacted - Fujitsu, IBM, Hewlett Packard, and Maxoptix. Typical disk seek times of 24 milliseconds, latency times of 12 milliseconds, and disk exchange times of 8 seconds were reported in the brochures obtained. However it is expected that access times will fall down in the near future to

about 10 millisecond when a direct overwrite feature is introduced. The storage capacity can be extended considerably to provide Terabyte ; of data by adding more SCSI controllers with their respective jukeboxes to the server. A 1.3 Gigabyte cartridge today costs approximately $80.

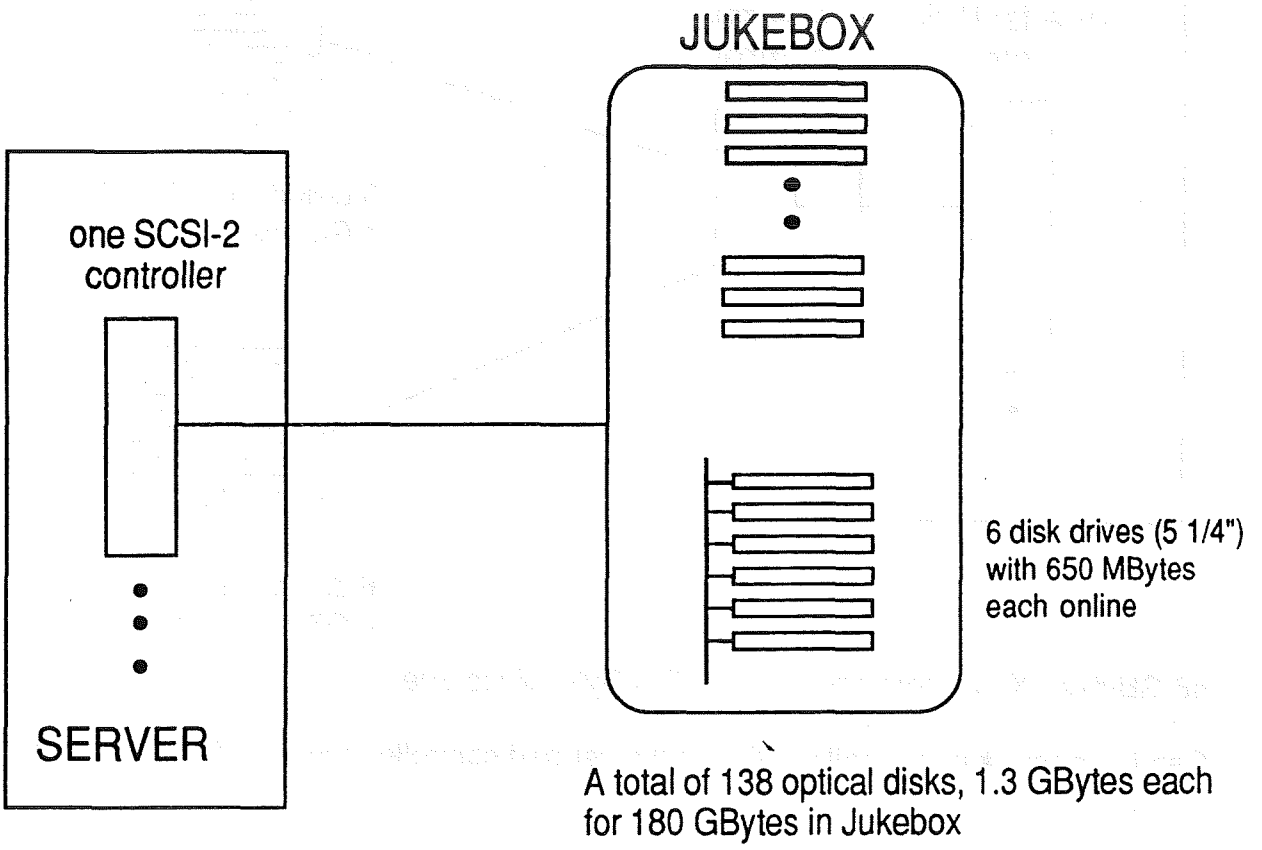


Figure 2-2. Jukebox Configuration

2.3 3-D OPTICAL MEMORIES

This type of memory belongs to the category of future technologies. It is predicted that a working prototype might be available in 3 years and as a commercially available product in 8 years.

At the Center for Molecular Electronics at Syracuse University, Dr. Robert Birge is developing a 3-D memory using a photoreactive protein called bacteriorhodopsin. Writing to this type of memory is accomplished by the absorption of two laser beams at different wavelengths. Reading is accomplished by detecting the light emitted by the protein when excited by a laser beam.

This type of memory can be used to implement write-once-read-many (WORM) devices or also

to implement read-write-erase devices. Other advantages include room temperature operation, inexpensive material, and parallel operation.

Dr. Birge was very helpful when asked about this technology and provided copies of quite a few of his publications that describe his ongoing work. These papers have been included with the separate document given to my NASA colleagues.

III

NEURAL NETWORK TECHNOLOGY

Neural networks are essentially a group of highly interconnected and relatively simple computational units as illustrated on Figure 3-1. Each of these computational units performs relatively simple processing of its inputs to produce a single output. The output of a unit is connected to the inputs of many other units through different weights.

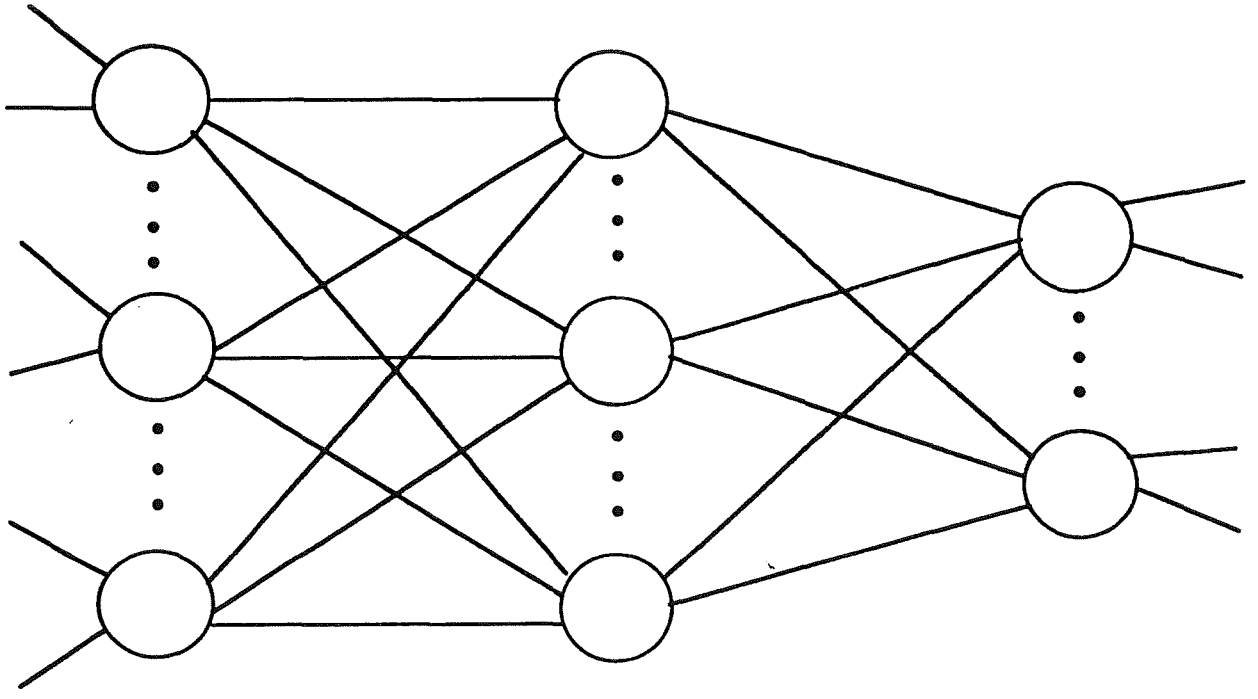


Figure 3-1. An Illustration of a Neural Network

Figure 3-2 shows a typical artificial neuron which adds all of its weighted inputs and uses a sigmoid output function to generate its output. Instead of a sigmoid function a threshold logic function is often used when binary functions are being implemented by neural networks.

3.1 WHY NEURAL NETWORKS ARE USEFUL

There are several reasons why neural networks have been used to solve a variety of problems. Probably the most significant capability of neural networks is that of learning previously unknown relationships directly from raw data. Indeed, neural networks are best suited for problems for which data exists that associates a set of inputs and outputs but for which no explicit relationships between them has been established. A typical example of this is the

automatic recognition of printed characters.

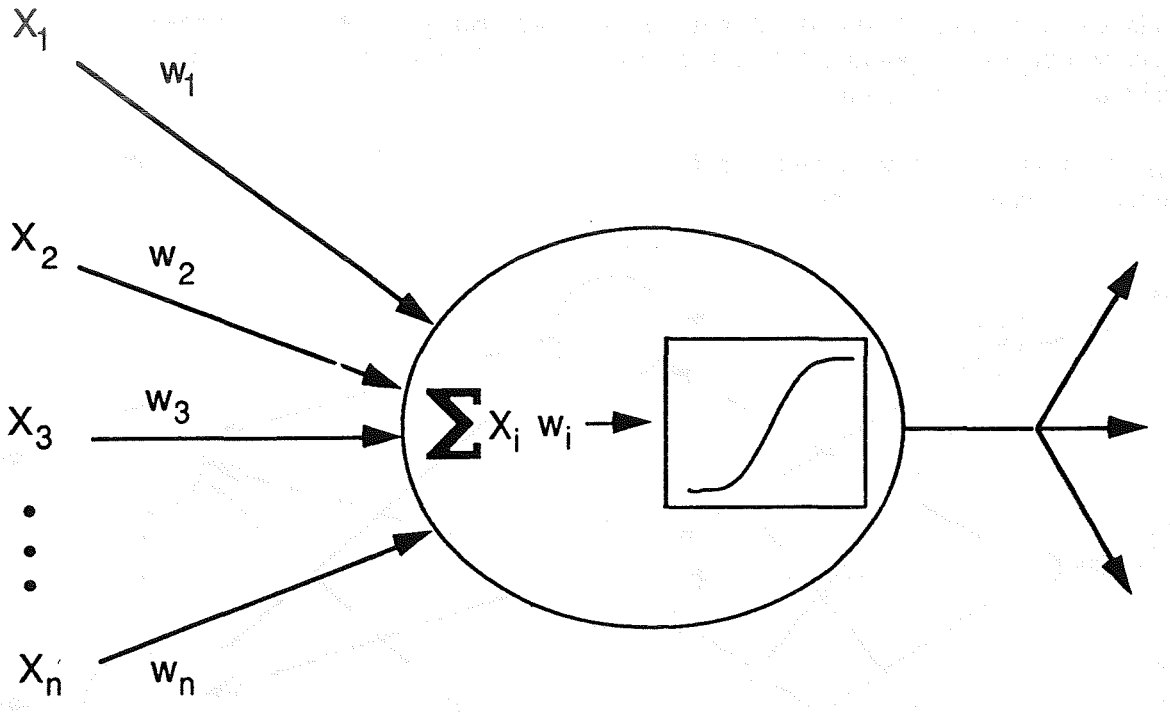


Figure 3-2. A Typical Artificial Neuron

Neural networks are capable of discovering nonlinear relationships from a training set and then apply these relationships to previously unseen data from that problem domain with very accurate results. This learning capability makes neural networks adaptable and very useful in environments where the relationships between input and output change over time.

Another important capability of neural networks is that they are fault tolerant. As individual neurons fail the performance of the entire network does not drop suddenly to zero but instead it slowly degrades in proportion to the number of failed units. This is in stark contrast to traditional algorithmic solutions where a single bit failure can have catastrophic results.

Finally, neural networks can fu ıction in parallel. The corresponding increase in speed that results from this can be used in applications requiring real-time responses.

3.2 TRAINING A TYPICAL NEURAL NETWORK

There are many types of neural networks. Among the ones most commonly described in the literature and included in the commercial software packages are: Multilayer Feedforward

networks, RCE networks, Radial-basis Function networks, Hopfield networks and Self-Organizing networks. Among these, 2-layer feedforward networks are probably the most popular because it has been shown that they are theoretically capable of implementing any association between input and output and because there is a well defined training method (Backpropagation) for them.

A typical 2-layer fully connected feedforward network is shown in Figure 3-3 with an n-dimensional input and m-dimensional output.

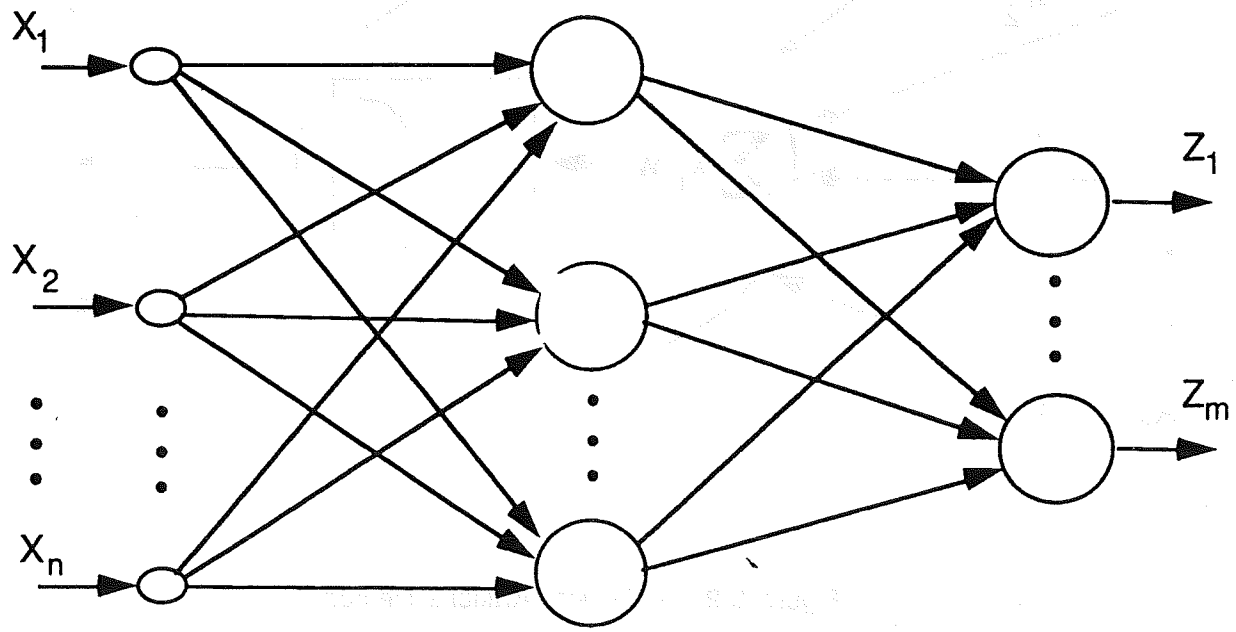


Figure 3-3. A 2-layer Feedforward Network

Training a network usually consists of finding the weight values so that previously known associations between input and output can be duplicated by the network. This implies that training requires a set of input data for which correct outputs are known. The learning that the network does through this method is called supervised learning. The correct weights can be found by an iterative procedure called back propagation. Training also normally requires manually changing the number of neurons in the hidden layer and sometimes increasing the number of layers.

An example of a training set of data is given in figure 3-4. The number of training examples required depends on the size of the network and the accuracy desired. Also a percentage of the training data collected (10 % is given as a guideline) is set aside for testing the network and is not used for training.

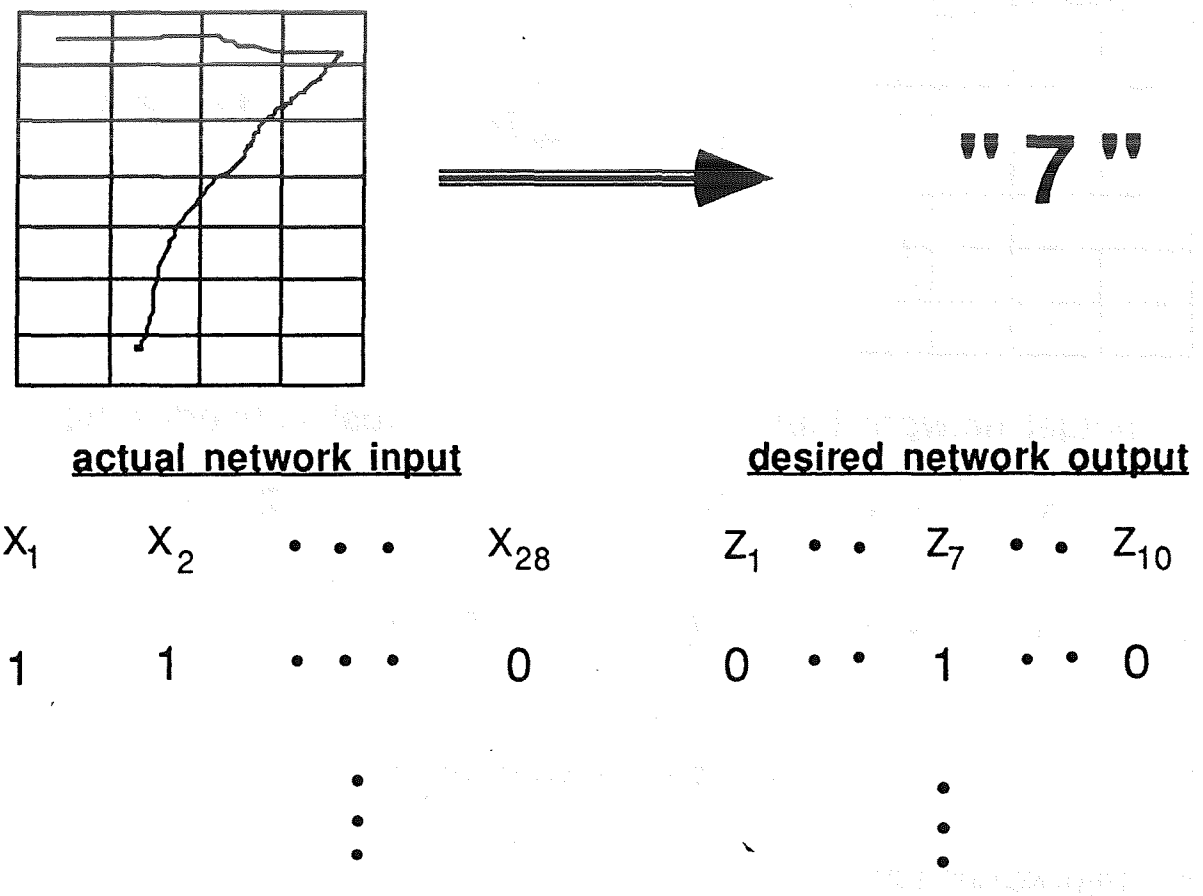


Figure 3-4. Training Data

Backpropagation uses the training examples to change the weights so that the error between the desired network output and the actual output is minimized. It is possible however to have a small training error but a large error when the test set is used. In many of these cases the problem is that the number of neurons in the hidden layer is too large. Thus the test set results and the training set results must be used in combination to arrive at the best network architecture - one that would perform well not only on the training set but also on the test set.

If the network has been trained properly, it will be able to correctly identify inputs that were not included in the training set. This is illustrated in figure 3-5. When this occurs the network has not memorized the training set but instead it has learned the basic relationships that exist in that problem domain between input and output and has stored those relationships in the weights of the network.

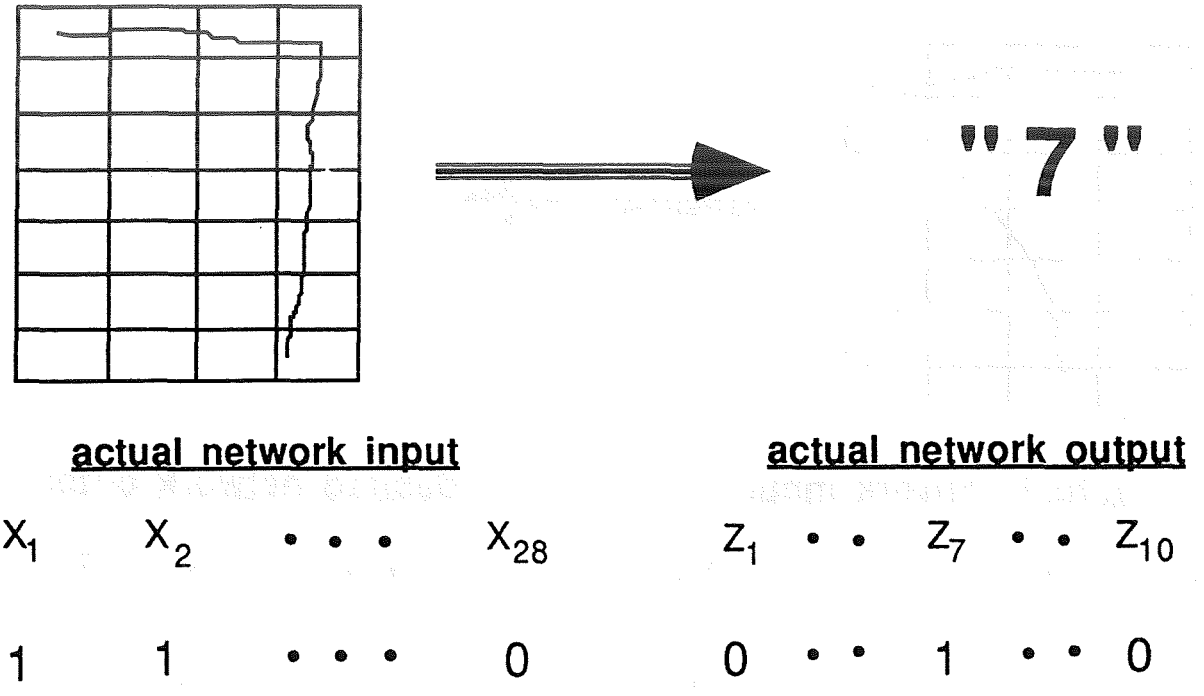


Figure 3-5. Network Response

3.3 TRAINING GUIDELINES

Guidelines for collecting network training data and for preparing the data for input to the network have been published in several books. The one by Jeannette Lawrence "Introduction to Neural Networks" is especially helpful on this topic.

When selecting a training set the following guidelines can be used:

- The number of training examples should be greater than the number of hidden neurons divided by the error criteria.
- Save about 10% of the training data for testing the network.
- Noise may be added to the original training data to increase the size of the training set.
- If implementing a network to classify data into different categories, include examples from each category. Include "border" examples for training and testing. If the network is to approximate a continuous function, include data throughout the entire range.
- Make sure that no contradictions exist in the training data. Also make sure that data that covers a long period of time has not been affected by extrenuous factors.

Before the data collected for training can actually be used by the network, it must be transformed into a number. If the data can be represented by binary values then this number will be 0 or 1 (sometimes the binary numbers used are -1 or 1). Network inputs that represent attributes of a problem that can take on symbolic values are usually represented by binary values as illustrated in Figure 3-6.

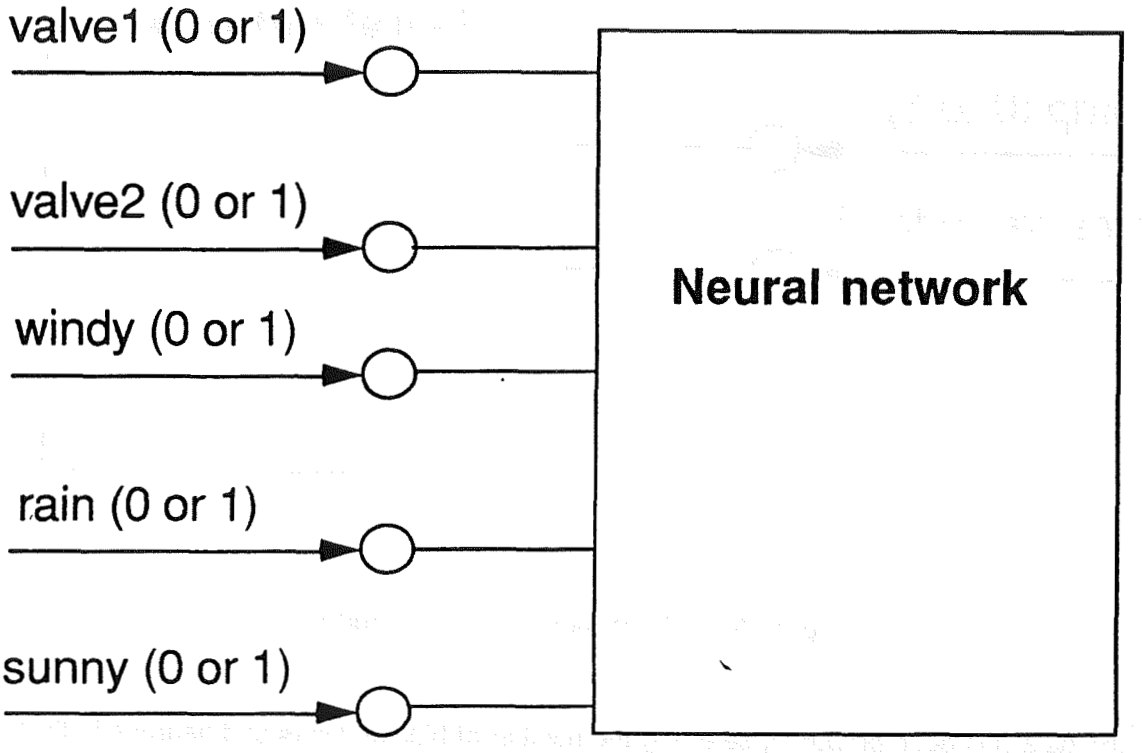


Figure 3-6. Binary Data Preparation

Network inputs that represent attributes of a problem that can take on numerical values are usually represented by continuous values between 0 and 1 as illustrated in Figure 3-7.

When the value range of a given attribute is very large, using the change recorded for that value instead of the actual value is better. This should be done only if it has been determined beforehand that it is the change in value that is significant for the problem at hand. Along the same lines, some data would need to be transformed using methods like Fourier's to reduce the dimensionality or to eliminate meaningless time variations in the original data.

If the goal of the neural network is to make a prediction based on what has happened in previous time periods, the network inputs should be set up as separate set of inputs for each of the time periods included in the data.

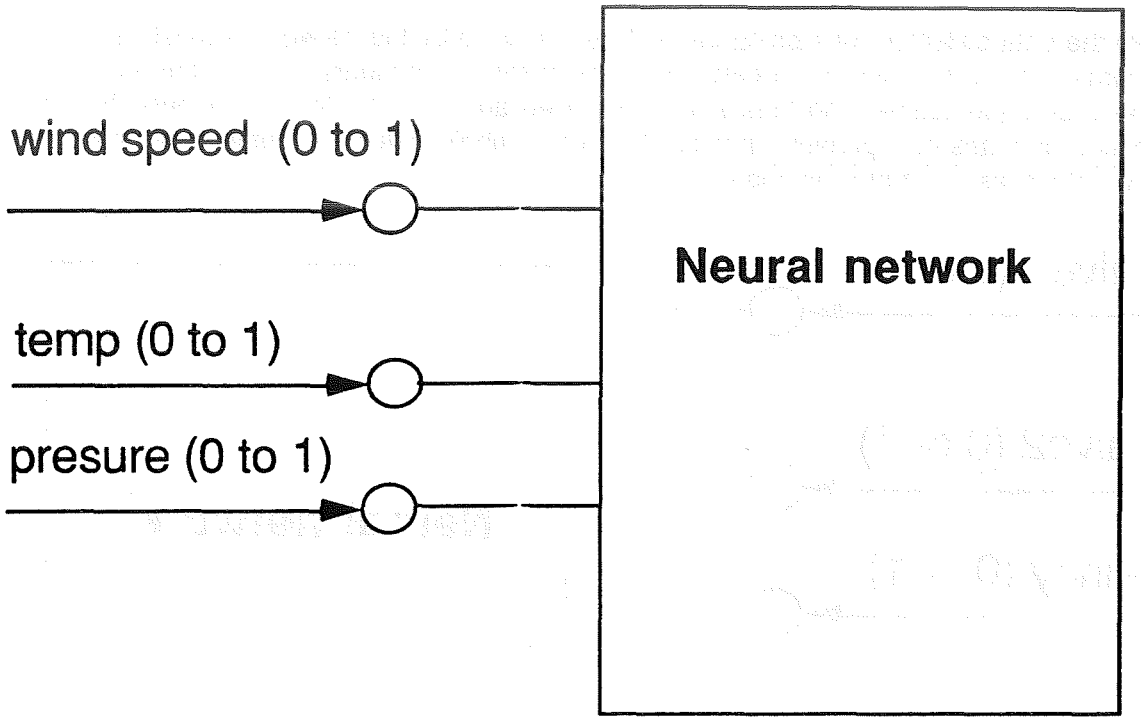


Figure 3-7. Continuous Data Preparation

After the data has been prepared, selecting the number of hidden layers and neurons in those layers is the next step in the training process. There are no clear guidelines on the number of hidden layers. Some suggest that one hidden layer is preferable because it is theoretically sufficient and improves training speed. Others suggest two hidden layers because the network architecture would fit the problem more effectively by allowing the first hidden layer to detect local features and then the second hidden layer to use these local features to detect global features.

The number of neurons in the hidden layers should be selected small at the start. After the network is trained it should then be tested. A few hidden neurons should then be added and the network retrained and retested. This process should be continued as long as the test performance continues to decrease. This should ensure that the network does not memorize the training set but that instead learns the general relationships that exist between input and output.

Other guidelines that are useful for training include the following: The learning rate should be smaller in the last layers than the front layers; initialization of the weights should be uniformly distributed within a small range; the training examples should be presented to the network in a random fashion and not according to categories.

3.4 NEURAL NETWORK SOFTWARE

There is a fair number of neural network simulation packages available. Some are commercial products from relatively small companies and others are available free of charge from university development labs. A representative sample of good packages from both categories are included below. The platforms on which they run, their list price, and a brief comment on their best distinguishing features are included.

3.4.1 COMMERCIAL NEURAL NETWORK SOFTWARE

- NeuralWorks Predict
  DOS, Windows, Macintosh $1995
  Automatic selection of best architecture
  Trained network is converted into C Fortran, or Visual Basic

- NeuralWorks Professional II/Plus
  PC/MAC $3995 - SUN $6995 - RS 6000 $7995
  SGI $7995 - HP $7995
  Very complete and powerful system with many network models
  C code generation

Both packages above are available from:

NeuralWare
Penn Center West Bldg IV
Pittsburgh, PA 15276-9910
412-7878222

- IBM Neural Network Utility
  Windows, OS2 $1500
  RS6000 $4995
  AS400 $550 to $9075 depending on the number of users
  Graphical interface, many network models, access to databases
  C code generation

  IBM
  800-3426672

- The Owl
  DOS, Windows $295
  UNIX Workstations $995
  Many network models

  HyperLogic
  1855 E. Valley Parkway, Suite 210
  Escondido, CA 92027
  619-7462765

- NeuroWindows
  DOS, Windows $369

- Neuroshell2
  DOS $195 - Windows $495
  Graphical interface and many network models

Both packages above are available from

Ward Systems Group, Inc
Executive Park West
5 Hillcrest Drive
Frederick MD 21702
301-6627950

- Brainmaker Pro
  DOS, Windows $795

California Scientific Software
10024 Newton Rd
Nevada City, CA 95959
800-2848112

3.4.2 FREE NEURAL NETWORK SOFTWARE

- PlaNet
  UNIX
  Nice graphical output, very good tutorial
  ftp://boulder.colorado.edu/pub/generic-sources/
  e-mail miyata@sccs.chukyo-u.ac.jp

- SNNS
  UNIX
  Graphical interface for creating and running network
  ftp://ftp.informatic.uni-stuttgart.de/pub/SNNS/

- Xerion and UTS
  UNIX
  Nice graphical output
  ftp://ftp.cs.toronto.edu/pub/xeron/
  e-mail xerion@ai.toronto.edu

- PDP
  UNIX and DOS (source code)
  Simple to run but primitive interface
  ftp://nic.funet.fi/pub/sci/neural/sims/

3.4.2 HARDWARE AVAILABLE

- Ni1000 digital chip
  ~$10000 for board and development software
  3.7 million transistors, 16.5 billion operations/sec
  256 inputs and 64 outputs
  Significant number of commercial applications

  Nestor, Inc.
  One Richmond Square
  Providence, RI 02906
  401-3319640

IV

ASYNCHRONOUS DATA TRANSFER TECHNOLOGY

Asynchronous transfer mode (ATM) is the latest technique developed for communication in computer networks. It is being promoted as the most effective way to increase the bandwidth of existing computer networks. It is also promoted as the communication technology that will allow the simultaneous transmission of audio, video, and data within the same network.

There are several reasons given for the predicted success of ATM. ATM, in contrast to ethernet or token-ring, uses a switching scheme that connects two nodes for the time needed to transfer data and it uses fixed size data packets. It also has theoretical speed capabilities on the order of gigabits per second. This makes ATM capable of transmitting audio and video signals which require large bandwidth and are very sensitive to the time and the sequence in which the information arrives. ATM is also independent of upper layer communication protocols and can be used with existing network architectures such as ethernet, and token ring providing a reasonable migration path for existing networks requiring increase bandwidth.

The following specific questions regarding ATM technology are of special interest to NASA:

- Is ATM being offered by a significant number of reputable vendors ?
- Have ATM standards been agreed upon ?
- Does ATM implement multicasting ?
- Which fault tolerant features, if any, are offered by ATM ?
- What interfaces are available for ATM ?

These questions are addressed below.

4.1 VENDORS OF ATM TECHNOLOGY

There are significant numbers of reputable vendors of ATM products. The picture that emerges from the different brochures and white technical papers from these vendors is that ATM technology will dominate the computer network market as existing networks try to increase their bandwith. The following vendors were contacted by phone or via the world wide web or both. Brochures and papers obtained from these vendors were placed in a binder and given to my NASA colleages.

- 3Com Corp.
  407-6611999 Ben Johnson
  http://www.3com.com
  3Com Corp
  2250 Lucien Way

Suite 100
Maitland, FL 32751

- AT&T
  201-6062978 Bill Price
  908-2212935 Christine Corliss
  301-6084596 Tim Russos
  http://www.att.com

- Cabletron Systems
  603-3329400
  http://www.ctron.com
  35 Industrial Way
  Rochester NH 03867

- Cisco Systems Inc.
  800-5536387
  http://www.cisco.com

- IBM
  800-4262255
  http://www.ibm.com

- Newbridge Networks Corp
  http://www.newbridge.com

- Whitetree Network Technologies
  415-8550855
  http://www.whitetree.com
  3200 Ash St.
  Palo Alto, CA 94306

In addittion, the University of New Hampshire's Interoperability lab tests ATM products from many of these vendors and is a good source of information about ATM. Another good source of information is the ATM forum, to which many of these vendors belong. It is an organization started in 1991 dedicated to promoting ATM and disseminating information about ATM. At present it has over 500 members and includes network equipment vendors, service providers, carriers, semiconductor manufacturers, and users.

- Interoperability Lab
  University of New Hampshire
  603-8620204 Ron Pashby
  http://www.iol.unh.edu

- ATM Forum
  http://www.atmforum.com

4.2 OTHER ATM ISSUES

It appears that 3 different sets of standards are being followed at the present time. The first one is RFC 1483 which includes bridging formats and routing protocols. The second one is RFC 1577 which includes IP over ATM. And the third one is LAN emulation. At this time it is not clear if any of these will eventually be accepted as the standard. The ability to multicast is provided by bridges and routers as well as by those products that provide IP over ATM. However ATM LAN emulators do not provide multicasting capabilities. The lack of clear standards makes it very unlikely that third party vendor hardware will be widely available.

It is possible today to purchase products which interface ethernet, token ring and FDDI networks to an ATM network. These products can process TCP/IP and IPX routing protocols and can convert shared media packets to ATM cells. There are also available products to connect workstations to an ATM network. These are interface cards which can convert the workstation data to an ATM cell.

It appears that at the present time most vendors are concentrating on offering ATM as a solution to the problem of limited bandwidth with today's ethernet and token ring based networks. It is rare to find an application which integrates audio, video, and data. Increasing data bandwidth seems to be the bulk of today's demand for ATM products.

However the cost of ATM products today is rather steep in comparison with other networking products. Network adapters cost between $2000 to $4000 and switches cost between $5000 and $10000. It is generally agreed that this high cost of ATM products as well as the lack of standards presents a serious obstacle to the rapid spread of ATM use.

V

EXPERT SYSTEMS TECHNOLOGY

Expert systems technology has been successfully used since the 1970's for a wide range of problems. It is considered today by most people to be a proven technology. At Kennedy Space Center quite a few projects based on this technology have been undertaken. Below is a list of recent or current projects at KSC funded through the Technology Programs and Commercialization Office. This list was provided by Carrie Parrish from the KSC Technology Transfer Office.

5.4.1 EXPERT SYSTEMS PROJECTS AT KSC

- Expert Systems for Operations Distributed Users (EXODUS)
  Contact: Jody Fluhr / TV-GDS-5 / 861-3788
  A platform of common utilities and services to facilitate the development, maintenance, and standardization of advanced software systems at KSC, including: data acquisition, recording and retrieval services; electronic documentation and measurement database services; and heterogeneous network communication services.

- Knowledge-based Autonomous Test Engineer (KATE)
  Contact: Peter Engrand / DM-ASD / 867-3770
  A generic, model-based reasoning shell for monitoring, fault detection and diagnosis of launch processing systems. The model-based system compares predicted performance with actual performance to determine system integrity.

- Vehicle Health Maintenance System (VHMS)
  Contact: Warren Lackie / TV-PEO-2 / 861-3968
  An expert system to monitor day-to-day vehicle operations (major power-on systems only) that will provide rapid detection and analysis of anomalies. KATE was used as the underlying model-based reasoning technology.

- Propulsion Advisory Tool (PAT)
  Contact: Bob Beil / TV-FSD-3 / 861-3944
  A Shuttle Main Propulsion System (MPS) advisory system that provides enhanced displays and plotting techniques, anomaly prediction, detection, warning and corrective action, trend analysis, and system diagnostics for use by the MPS console operators.

- Reasoning Based on Intelligent Computers and Networking (RUBICON)
  Contact: Scott Wilson / TV-GDS-4 / 861-3846
  A generic shell for KSC expert systems' technologies that provide telemetry data, database access, common utility functions, and a design standard for future expert systems.

- Test Management Plan Automation
  Contact: Rudy Tench / LSO-217 / 861-7435
  A system that applies available project management tools and automation techniques to the process of producing a control sequence which is a textual interpretation of a detailed

schedule for all of the work instructions that support Shuttle processing.

- Advanced Shuttle Scheduling Technology, Ground Processing Scheduling System (GPSS)
  Contact: Nicole Passonno / TV-PED-2 / 861-5434
  An interactive, intelligent scheduling tool for Shuttle ground processing based on vehicle configuration, constraints, and resources.

- Advanced Data Management System (ADMS), Structured Surveillance System
  Contact: Randy Tilley / RM-INT / 867-2020
  A system to be used for the collection, manipulation, and dissemination of information collected by NASA Quality Inspectors for the Structure Surveillance program.

- SR&QA Portable Data Collection (PDC)
  Contact: Randy Tilley / RM-INT / 867-2020
  A data collection system that provides accurate, near-real-time status of work in progress, as well as work completed. The technology may involve the use of pen-based computers, electronic stamps, and the Wireless Information Network (WIN).

- Smart O&M Manual Development Project
  Contact: Carolyn McCrary / DL-ICD-I / 867-4449
  A system that used CD-ROM storage, search and retrieval capabilities for system engineer's documentation, such as Operations and Maintenance (O&M) manuals, to enable a decrease in storage requirements and access time to information.

- Mapping, Analysis and Planning System (MAPS)
  Contact: Burton Summerfield / MD-MED / 867-4237
  A centralized information database system that provides KSC operational elements with environmental compliance, management and impact assessment information through the use of a knowledge-based decision support system.

- Automated Database Design from Natural Language Input
  Contact: Carl Delaune / CG-ISO-1 / 867-3526
  The implementation of natural language understanding, knowledge acquisition, knowledge representation, and problem solving methodologies as an interface for the design and development of databases through natural language input.

- Electronic Performance Support System (EPSS) Research and Applications
  Contact: Dick Davis / DE-TPO / 867-2780
  An intelligent, interactive multimedia database system that provides quick assistance and information without the need for user training. EPSS may incorporate all forms of multimedia delivery, as well as artificial intelligence techniques such as expert systems and natural language recognition.

- Worksite Thunderstorm Prediction System
  Contact: Frank Merceret / TM-LLP-3 / 867-2666
  A mesoscale modeling system designed to provide accurate forecasts of specific thunderstorm-related phenomena such as precipitation and high winds thereby reducing

downtime due to false weather advisories and alerts, hazardous weather events occuring without warning, and unnecessary restrictive weather-based flights rules for manned and unmanned missions.

5.4.2 COMMERCIAL EXPERT SYSTEMS DEVELOPMENT SOFTWARE

A number of reliable software vendors were identified and their brochures provided to my NASA colleages. As with neural network software, hooks to databases are also available with expert system development software on a variety of operating systems and platforms making integration within an advanced launch processing center very feasible.

- ART
  Windows, UNIX

  Brightware Inc.
  415-8999070
  101 Rowland Way, Suite 310
  Novato, CA 94945

- G2
  Windows, Windows NT, UNIX, VMS

  Gensym Corp.
  617-5472500
  125 CambridgePark Drive
  Cambridge, Massachusetts 02140

- Kappa
  Windows, UNIX $9995

  Intellicorp
  415-9655500
  1975 El Camino Real West, Suite 101
  Mountain View, CA 94040-2216

- M.4
  Windows, $995

  Teknowledge
  415-424-0500
  1810 Embarcadero Rd.
  Palo Alto, CA 94303

- Nexpert Object
  Windows, Windows NT, OS/2, Macintosh, UNIX

Neuron Data
415-3214488
156 University Avenue
Palo Alto, CA 94301

- RTworks
  UNIX, VMS

  Talarian Corp.
  415-9658050
  444 Castro St. Suite 140
  Mountain View, CA 94041

VI

CONCLUSIONS

Advanced technology available today can significantly improve the design of a new shuttle launch processing center. The following conclusions can be made regarding the technologies that were reviewed within the 10 week time period asigned to this project.

The latest high capacity data storage technology today, as reported by a group of reputable vendors, consists of disk arrays and magneto optical disks. Each is capable of storing more than the 1.3 Terabytes of data that it is estimated constitute today NASA's requirements for launch related data. Disk arrays are more expensive than magneto optical disks but they provide faster access to data. A 180 Gigabyte disk array would cost $80,000 today not including the server, while an optical jukebox of similar capacity would cost $50,000. Disk seek times of 10 milliseconds and latency times of 5 milliseconds are typical while magneto optical disk seek times of 24 milliseconds and latency times of 12 milliseconds are the norm - although it is expected that the access speed for optical disks will decrease within a year or two. In addition, disk arrays make all of the data available online, while a jukebox of magneto optical disks may provide only about 4 gigabytes of online data. Furthermore, it takes 8 seconds to load an optical disk within the jukebox for data not found online. It is quite possible that a storage system that combines both of these technologies, i.e. disk arrays for critical data that must be accessed quickly and optical disks for the rest, will provide a good solution for storing data in a new launch processing center.

Neural networks software has been successfully used for a variety of problems. A number of software vendors can provide reliable and effective development packages for neural network solutions which can execute on DOS, Windows, Mac, or Unix environments. Reasonable hooks from these products to databases are provided so that integration with a data storage system should be feasible. A commercially available hardware implementation of neural networks that could be used for speed critical applications also exists.

It is important that the selection of launch related problems for the application of neural network technology be done carefully. The person or team involved should be thouroughly familiar with the problem domain and with neural networks. The problem should be one where the relationships between cause and effect have not been explicitly identified but where there is sufficient historical data from which to learn these relationships. Neural networks can be retrained quickly to learn new relationships and therefore can also be effective for problems that frequently change.

Asynchronous transfer mode (ATM) is the latest technique developed for communication in computer networks and is expected to become the dominant networking technology in the future. Its main advantages are a transmission speed on the order of gigabits per second and a dedicated switched connection between sender and receiver nodes. These can allow the simultaneous transmission of audio, video, and data within the same network. Presently the main disadvantages of ATM are the lack of standards and the high cost of ATM products.

There are significant numbers of reputable vendors of ATM products. Brochures and papers

obtained from these vendors were placed in a binder and given to my NASA colleages. It is possible today to purchase products which interface ethernet, token ring and FDDI networks to an ATM network as well as to purchase interface cards that connect workstations to ATM networks. It appears that at the present time most vendors are concentrating on offering ATM as a solution to the problem of limited bandwidth with existing networks. It is rare to find an application which integrates audio, video, and data. Increasing data bandwidth seems to be the bulk of today's ATM applications.

Expert systems technology has been successfully used since the 1970's for a wide range of problems. It is considered today by most people to be a proven technology. At Kennedy Space Center quite a few projects based on this technology have been undertaken. A number of reliable software vendors was identified and their brochures provided to my NASA colleages. As with neural network software, hooks to databases are also available with expert system development software. They run on a variety of operating systems and platforms making integration within an advanced launch processing center very feasible.

1995 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

S16-51
7756
P. 20

AN ANALYSIS OF ALTERNATIVE TECHNOLOGIES FOR THE REMOVAL OF ETHYLENE FROM THE CELSS BIOMASS PRODUCTION CHAMBER

Dr. Allen L. Rakow
Associate Professor
Department of Chemical and Bioresource Engineering
Colorado State University
Fort Collins, Colorado

KSC Colleague - John Sager
Life Sciences

Contract Number NASA-NGT-60002
Supplement 19

August 9, 1995

ACKNOWLEDGEMENTS

The CELSS group was once again supportive of my task. In particular, I am grateful for the help I received from Ray Wheeler, Gary Stutte, Russ Fortson, John Sager, Barry Finger, Dick Strayer, and Barbara Peterson. In addition, Ray Hosler, Kari Stiles, and Calleen Coiner were instrumental in making my summer stay an enjoyable and productive experience.

ABSTRACT

A variety of technologies were analyzed for their potential to remove ethylene from the CELSS Biomass Production Chamber (BPC). During crop production (e.g., lettuce, wheat, soybean, potato) in the BPC ethylene can accumulate in the airspace and subsequently affect plant viability. The chief source of ethylene is the plants themselves which reside in plastic trays containing nutrient solution. The main sink for ethylene is chamber leakage. The removal technology can be employed when deleterious levels (e.g., 50 ppb for potato) of ethylene are exceeded in the BPC and perhaps to optimize the plant growth process once a better understanding is developed of the relationship between exogenous ethylene concentration and plant growth. The technologies examined were catalytic oxidation, molecular sieve, cryotrapping, permanganate absorption, and UV degradation. Upon analysis, permanganate was chosen as the most suitable method. Experimental data for ethylene removal by permanganate during potato production was analyzed in order to design a system for installation in the BPC air duct. In addition, an analysis of the impact on ethylene concentration in the BPC of integrating the Breadboard Scale Aerobic Bioreactor (BSAB) with the BPC was performed. The result indicates that this unit has no significant effect on the ethylene material balance as a source or sink.

SUMMARY

Several technologies were examined in order to find a suitable method for removing ethylene from the CELSS BPC when detrimental levels are present. For example, during a normal potato growth cycle ethylene levels are usually in the 20 -40 ppb range, therefore, when levels reach the 50-60 ppb range a removal process is initiated. The process consists of placing 200 grams of a material containing potassium permanganate near the air supply duct inside the chamber after which the ethylene concentration is logged on a portable gas chromatograph. When normal levels are reached the permanganate is removed from the chamber. This currently takes several days and is unsafe because the material can be hazardous for both humans and plants. An alternative approach would include the placement of a removal system in the air duct surrounding the chamber so as to eliminate entry into the chamber. Such a removal system would have to meet several design constraints including the avoidance of significant upset to chamber temperature, pressure, humidity, and carbon dioxide levels. In addition it would have to be safe, reliable, and cost efficient. For this purpose four different technologies were analyzed for potential application along with consideration for modification of the existing technique. The four methods examined were catalytic oxidation, molecular sieve, cryotrapping, and UV degradation. None of these methods were acceptable, in part due to their propensity to upset conditions in the BPC. As a result potasium permanganate was explored in further detail beginning with an analysis of ethylene removal during potato growth in May, 1995. Using the material balance model developed by the author in 1994, it was determined that the removal rate is about one half a ppb per hour. This rate was then used for scale up to larger amounts removed as well as faster rate of removal. Four kilograms of permanganate material might be adequate to bring ethylene down to a safe level within an eight hour time period for most conditions encountered in the BPC. In addition an analysis was done to determine whether or not the Breadbord Scale Aerobic Bioreactor contributes in a significant manner to the total ethylene material balance. An upper limit of less than one ppb per day coming from the BSAB was calculated which can safely be neglected in the total balance equation. Finally, recommendations for future experiments were made in an effort to establish whether plants respond autocatalytically during the normal growth phase, i.e., whether exogenous levels of ethylene promote the production of ethylene. Should this be the case, then a control system for ethylene might be employed to optimize the plant growth process.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

I. Introduction

When different crops are produced in the CELSS BPC, the plant hormone, ethylene is produced as well, albeit at different rates throughout the plant cycle. When the production rate of ethylene exceeds the chamber air leakage rate ethylene accumulates in the BPC at parts per billion levels. Conversely, when the leakage rate is greater than the production rate the chamber concentration declines. In a closed and sealed BPC there is always the potential to generate a deleterious level of ethylene. In the case of potatoes 50 ppb might be problematic, therefore, a removal procedure is currently initiated when the concentration reaches this level in order to avoid possible effects on crop viability.

The current removal method involves the placement of approximately 200 grams of pellets containing potassium permanganate inside the chamber near the air supply duct. The permanganate is hazardous (to both humans and plants) and if spillage occurs, especially into the circulating nutrient solution, the plants could be affected. A better, and probably more effective way of removing ethylene when necessary, would involve a system remote to the BPC, i.e., one that could be installed in the surrounding air duct and activated when needed. For this purpose several different technologies were analyzed in the hope of replacing the current procedure.

The following technologies were considered as alternates to the current method:

1) catalytic oxidation
2) molecular sieve
3) cryotrapping
4) UV degradation

Each technology was analyzed with respect to design constraints imposed on the system and further analysis of a workable method was performed.

In addition, an analysis of the Breadboard Scale Aerobic Bioreactor's contribution to ethylene concentration in the BPC was made in a similar manner to the analysis of sources and sinks for a material balance model established by the author in 1994.

II. Ethylene Dynamics

In 1994 the author established a material balance model for ethylene in the BPC air space. A primary assumption in the model is that the ethylene concentration in the air space is uniform throughout the BPC, i.e., any ethylene that is of-gassed from the plants is instantaneously distributed throughout the entire airspace. This assumption is premised on the hefty air circulation rate in the BPC (~4 volume changes per minute). It was determined that on a daily average basis there is only one significant source (plants) and one significant sink (leakage) for the determination of the accumulation rate. The accumulation rate was obtained from the first derivative of the concentration curve and the leakage rate was estimated independently from carbon dioxide decay experiments when no plants were in the BPC. The plant production rate was obtained by adding the leakage rate to the accumulation rate and was in agreement with other studies on ethylene production.

Plants produce ethylene at different rates at different stages in the production cycle and the process is not fully understood. One aspect that is particularly important to the BPC is whether or not the plant production rate is dependent on the exogenous concentration since the rate increases with increasing concentration until perhaps some trigger (concentration?) causes the production rate to fall below the leakage rate and the chamber concentration begins to decline. In fact there is evidence (1) of autocatalytic behavior (autoinhibition is also possible) for some plants. On the other hand the increased production may be an endogenous response and may correlate with some plant growth characteristic such as leaf area, etc. Of course to make matters more complicated, both mechanisms could be at play at different times or even at the same time. If only one mechanism is at play it could be determined by dosing the chamber or leaking the chamber during normal plant growth and measuring the ethylene concentration for several hours afterwards. If the rate of production increases in the dosing experiment or decreases in the leakage experiment then the exogenous mechanism is at work. Of course, the dosing experiment would have to be carried out very carefully (small, but significant amounts) so as not to trigger any other response mechanism. If the exogenous mechanism prevails it is possible that adjusting the exogenous concentration could optimize biomass production.

In either case it is important to avoid abnormally high levels of ethylene in the BPC since there are a variety of negative effects that can ensue. In this regard it is imperative to have a reliable removal system.

If exogenous concentration drives production it is most likely a first order process since both accumulation and leakage are first order and a mechanism might be the following:

Receptor + Ethylene + Ethylene Forming Enzyme(EFE) > Ethylene + EFE* +Receptor

EFE* + 1-aminocyclopropane-1-carboxylic acid (ACC) > Ethylene + EFE

Net Reaction: Ethylene + ACC > 2 Ethylene *= activated enzyme

The reaction rate, r, =k [Ethylene] [ACC] or if ACC is in excess: r = k [Ethylene]

III. Removal System Design Constraints

The removal system will be located in the air circulation duct surrounding the BPC. The air circulation rate is 400,000 liters per minute so pressure drop should not be a problem; nevertheless, the removal system must not significantly affect the air circulation rate. This high flow rate is advantageous to any removal system since it will most likely eliminate any mass transfer resistance, i.e., the ethylene in the gas phase will be transported very rapidly to the surface of any removal system.

A very important constraint is that there be no significant upset to chamber conditions. The following conditions need to be maintained:

a) Temperature~ 20-25 C
b) Pressure ~ 1 atm
c) Relative Humidity ~ 70-90%
d) Carbon Dioxide Concentration ~ 1200ppm

The removal system should be able to reduce the ethylene concentration at a high enough rate to bring the concentration down to a normal level within an eight to ten hour period. Of course the system should be safe, reliable, and preferably low in cost.

Because of the high air flow rates in the BPC air duct it is important to have a removal system which , if porous, maintains its structure under flow. If the material collapses or gets compressed or altered such that the surface area is reduced this could have an effect on the removal efficiency.

IV. Candidate Technologies for Ethylene Removal

Several technologies were examined in an effort to find a suitable method for lowering ethylene concentration in the BPC when levels become problematic for a particular crop. With the exception of cryotrapping, each method has been used to scrub ethylene and literature data is available. First and foremost the technology was analyzed with respect to the design constraints. In most cases the literature data was obtained for small scale laboratory equipment and/or higher ethylene concentrations than those found in the BPC. Nevertheless an attempt was made to extrapolate to the BPC. Each method is presented in detail below.

4.1 Catalytic Oxidation

In a 1978 study Eastwell et al. (2) flowed a compressed air stream containing 20 ppb of ethylene and other hydrocarbons through a 45 cm long , 1.27 cm diameter tube containing platinum on asbestos fibers (5 w/w%) at 800 ml/min. The temperature was maintained at 650 C and all the ethylene was oxidized. The corresponding length of similar catalyst material upon scaleup to the BPC air duct would be 24 meters. This amount of catalyst is much higher than what is most likely required for two reasons. Firstly, a much higher weight fraction of platinum on a surface could be employed and secondly, the high concentration of the other hydrocarbons (ppm range) in Eastwell's study would not be present in the BPC. On the other hand the high temperature would preclude this technique in the BPC because it would cause a major upset.

Abeles (1) describes a Swingtherm catalytic converter which handles 10,000 liters of air (containing ppb levels of ethylene) per liter of catalyst per hour. When scaled to the BPC air duct this would require 3-4 meters of catalyst, a rather large amount. Although this system operates at much lower temperature (~200 C) it would still be prohibitive due to BPC upset capability. This technology could be employed outside the BPC as a free standing, portable unit which could draw the duct air as a bypass stream; however, the effluent to be returned to the BPC would have to be cooled and cost would be a definite factor.

Another problem with catalytic oxidation is that most, if not all of the ethylene would be removed (a finite amount of ethylene is probably good for crop production) if the system were within the air duct at current flow rates. Thus, the bypass technology (lower flow rates than in the air duct) might be the preferred approach with regard to the achievement of an optimum level of ethylene. On the other hand it will be difficult to match any arrangement with a desired level of reduction (a variable target).

4.2 Molecular Sieve

Jiang et al. recently flowed (1cc/min) a mixture of ethylene (8.5 mole%) and ethane through a small tube filled with 2-6 grams of Linde 5A Zeolite material at a

temperature of 30 degrees C and a pressure of 1 atm. The ethylene adsorption rate in this experiment was 0.000005 gmoles/min.

It would be difficult to size a molecular sieve for the BPC air duct based on Jiang's data because the concentration driving force and Reynolds number of the flow are so different. Also, the sieve will only be effective when most, if not all, of the water in the air stream is removed, i.e., the manufacturer recommends bone dry air. To make matters worse even carbon dioxide can reduce the efficiency of the sieve.

A thermal cycle would be necessary to restore the sieve after adsorption has taken place, i.e., the sieve adsorbs at low temperature and desorbs at higher temperature. This could be done when the sieve is saturated with ethylene using a swing system.

For the small amount of ethylene that needs to be removed in the BPC (~.001 gmoles) only 1 Kg of material is necessary (based on a scaleup of Jiang's data) and the cost is nominal. However, the water removal pre-stage would cause a major upset to the BPC unless water could be replenished rapidly downstream from the sieve. This would be difficult to accomplish without installing new humidification capability.

4.3 Cryotrapping

This technology would involve a countercurrent heat exchanger with the BPC air on one side and liquid nitrogen (boiling point of -320 F) on the other. The trace organic of interest (in this case ethylene) would desublimate and freeze out on the surface of the exchanger. There are two reasons why this method is not suitable for the BPC. Obviously there would be a major upset to the BPC which would require a restoration step downstream. Secondly, ethylene is too volatile to capture at its low concentration in the BPC. Based on equilibrium vapor pressure data (4) the equivalent saturation pressure in the BPC at liquid nitrogen temperature is 395 ppb. Since the problematic concentrations of ethylene in the BPC are in the 50-150 ppb range it would be impossible to freeze out the BPC ethylene.

Based on some preliminary data on ethylene concentrations in the shuttle(5) it is entirely possible that ethylene concentrations in human compartments in space could reach higher than 395 ppb even though it is not clear what the source of ethylene was in the preliminary data.

Cryotrapping might be a desirable technology to employ for lunar and martian stations since it could be used to control water and carbon dioxide levels as well as some trace contaminants by taking advantage of local low temperature conditions.

4.4 UV Degradation

The removal of ethylene by ultra violet radiation is accomplished with germicidal lamps at two different wavelengths, 184 and 254nm (6,7,8). The lower wavelength produces atomic oxygen and ozone, while the second generates molecular oxygen from the ozone produced by the first lamp. It is the atomic oxygen which reacts with ethylene. Unfortunately (6) not all of the ozone is reduced and the residual amount can be as high as 1000ppb. This level can be problematic to BPC crops which are sensitive at 100ppb(9). Otherwise this could be a very inexpensive and simple method for removing ethylene. Shorter and Scott flowed air containing 1200ppb of ethylene from a 600 liter chamber at 60 liters per minute through a 100 liter box containing only one UV lamp. Ethylene concentration was reduced to less than 20ppb in about 20 hours. The slope of the concentration curve is very small as the concentration decreases below 100ppb so the removal rate if applied to BPC concentrations would be very small. Therefore, a much greater number of lamps would be necessary to make this method practical. In any event, until a reliable UV method is developed where ozone is maintained below 100ppb while reducing ethylene, this technology is inapplicable. There is brief mention of a new method which produces ozone which in turn is reduced by catalysis (10), however, no data is given regarding ozone reduction. As more information is published on ambient catalysis it might very well become a viable technology.

4.5 Permanganate Absorption

Ethylene reacts with potassium permanganate and water to form ethylene glycol, manganese dioxide, and potassium hydroxide. The stoichiometry is the following:

$$3CH2{=}CH2 + 2KMnO4 + 4H2O = 3C2H6O2 + 2MnO2 + 2KOH$$

Heat and acid promote the further oxidation of the glycol. $KMnO_4$ absorbents have been used successfully to reduce ethylene levels in growth chambers and greenhouses. According to Abeles (1) the chief disadvantages are the need for replenishment (irreversible reaction) and safe disposal. However, the fact that the reaction is irreversible could be advantageous in the BPC air duct simply because the ethylene would be converted and there would be no need, as in a regenerable system, for ethylene disposal. Nor would there be any chance of re-introducing ethylene to the BPC.

The permanganate is bound to an activated alumina substrate and is initially purple in color. Upon reaction the color changes to brown thereby affording a relatively easy means of determining when the material is spent.

Lidster et al. (11) evaluated $KMnO_4$ as an ethylene remover from apple storages and found the highest removal when the material was arranged in a tubular configuration. The variables that they found to affect ethylene removal were temperature , relative humidity, particle size, pH, and air flow rate. High temperature increased removal

removal whereas high humidity decreased removal. In some cases alkalinity increased removal. Smaller particles and lower flow rates (packing factors?) increased removal.

Permanganate is the current method for ethylene removal in the BPC. The material currently comes in bulk pellet form so it is manually transferred to holding containers (air through) which are placed in the BPC when needed. Because humidity is a factor, the pellets are placed near the air supply prior to the plant trays. Unfortunately, spillage can sometimes occur and permanganate is hazardous to the plants. Also, the chamber has to be opened in order to bring in/remove fresh or spent material. This is an undesirable mode of control, therefore, a method which can operate remote to the plant chamber is desirable. In this regard Air Repair of Stafford, Texas manufactures permanganate in a tubular form consisting of a polypropylene net which houses robust pellets. The pellets are 3.5-5.0% KMnO4 by weight and contain 15-20% water and are alkaline in pH.

Summary: Of all the technologies considered only the permanganate has the potential (with modification) to meet the design constraints. Therfore, further analysis and scaling was performed for permanganate.

V. Permanganate Removal Experiment

At the beginning of May of this year a potato crop cycle was started in the BPC. About the 15th of the month the ethylene level in the upper chamber increased to about 60 ppb, therefore, 200 grams of pellets containing permanganate were placed in the chamber. The permanganate was removed on the 30th of May after the concentration had dropped to the 15-30 ppb range as measured by a portable gas chromatograph. In order to determine the effectiveness of the permanganate it is imperative to know precisely how much ethylene is lost to the outside by leakage. During this two week time period the chamber was entered on several occasions making it very difficult to estimate the true chamber leakage rate on an hourly or even daily average basis. The reason for this difficulty is the fact that each person entering the chamber does so in a different manner, i.e., some take more time than others to make their entry and/or exit and some close the door completely without sealing the chamber whereas others may leave the door ajar. Fortunately, there was a three day continuous period during the time that the permanganate was employed where the chamber was in a closed and sealed state. In this mode the leakage rate is known to be 10% of the chamber volume per day. The concentration of ethylene during this three day period is plotted in Figure 1 and although the data is highly scattered (this scatter could also be a diurnal ethylene pattern) it appears to exhibit a downward trend. In fact a linear fit to this data has a correlation coefficient of 0.5. Keep in mind that the sampling method involves a single small port through the wall of the chamber which has a volume of 56,000 liters. In other words there could very well be a sizeable boundary layer near the wall where mass transfer of ethylene produced by the plants is significantly different from the inner part of the chamber. Velocity mapping of the chamber might provide some explanation for the variability of the concentration. However, it is the downward trend that is most important for the computation of the removal rate.

VI. Permanganate Removal Rate

The first step in specifying a permanganate removal system for the BPC is the determination of the current removal rate which can be obtained from a material balance model for ethylene (12). For any time period when removal is occurring the balance is the following: $A = P - L - R$, where A is the accumulation rate in ppb of ethylene per hour, P is the plant production rate in ppb/hr, L is the chamber leakage rate of ethylene also in ppb per hour and R , in ppb/hr, is the removal rate. Since the concentration of ethylene is declining during the three day experimental period, the accumulation rate is negative, i.e., deccumulation occurs and is obtained from the first derivative of the concentration verses time curve which in this instance is a constant (.18 ppb/hr) throughout the concentration decline. The leakage rate, L, for a closed, sealed chamber is 10% of the average daily chamber concentration per day. On an hourly basis this would be 0.1/24 x C(t), where C is ppb and t is in hours. For this experiment the linear fit for C(t) is $C = 34.4 - .18\ t$; however, since the other rates in the material balance equation are average rates, C is taken to be an average value for the three day period (28 ppb). The production rate, P, is obtained from a material balance on potatoes in the BPC for day 23+ in the cycle (12) and is taken as an average value (.37 ppb/hr) for the three day period. The result is the following:

$$-.18 = .37 - .1/24\ (28) - R$$

Or R, the removal rate, is equal to .43 ppb/hr.

Another important rate is the speed at which the permanganate gets depleted. As an example, if you remove 20 ppb by permanganate in a drawdown mode then 5 mg of material are exhausted : 20 ppb x 56,000 liters x 1/22.4 gmoles/liter x 2 moles permangante/ 3 moles ethylene x 158g/gmole permanganate = .005 grams exhausted per drawdown. If a canister of absorbent material contains 200 grams of potassium permanganate then 200/.005 = 40,000 drawdowns per canister is a best case scenario(other volatile organic compounds in the BPC can react with permanganate). For a crop which has a higher production rate of ethylene or, if faster drawdown is desirable, additional canisters could be employed. However, without knowledge of the exact amount of permanganate in the material employed, the degree of competition from other compounds that react with permanganate, the long term stability of the porous structure, the effect of moisture on pellet characteristics and other factors it is difficult to predict drawdown capacity. Scaleup will also be difficult without knowledge of the different resistances to mass transfer of ethylene to the pemanganate surface in comparison with reaction rate.

VII. BSAB Ethylene Analysis

A Breadboard-Scale Aerobic Bioreactor has been integrated into the BPC loop in an attempt to achieve a closed cycle system. A mixed population of microbes digest plant waste in order to generate nutrient solution for the plants. The water for the 120 liter bioreactor is obtained from BPC condensate. Oxygen for the microbes derives from BPC air. A .75 inch diameter line from the BPC air duct feeds a compressor which sparges the air into the bioreactor. The total volume of the BSAB air loop is under 100 liters (negligible in comparison to the BPC's 113,000 liters).

Microbes have the potential to produce and/or consume ethylene, therefore, it is important to investigate what role they play as a source or sink in the total ethylene material balance. According to Abeles (1) ethylene does not appear to play a role in bacterial growth and development. Soil-borne species such as Mycobacterium paraffinicium can grow on ethylene whereas ethylene-producing bacteria might promote plant senescence through microbial attack to facilitate bacterial growth. Certain strains of Pseudomonas solanacearum associated with the premature ripening of bananas produce ethylene. In fact this species produces one ten billionth of a mole of ethylene per minute for every 10 billion bacteria present. The growth rate of these bacteria is greater than one billion cells per hour and is higher than BSAB populations, nevertheless, it could serve as a good upper bound for an estimate of BSAB ethylene contribution to the BPC.

Current bacterial counts for the BSAB are around one billion per ml. Assuming that the BSAB microorganisms produce ethylene at the same rate as the Pseudomonas salanacearum the BSAB production rate would be the following:

10^{-10} gmoles ethylene/min/10^{10} bacteria x 22.4 liters/ gmole x chamber/56,000 liters x 10^{9} parts/ billion parts x 120 liters x 1000 ml/ l x 10^{9} bacteria/ ml = .0005 ppb/min = .72 ppb/day

Given that potatoes produce ethylene at 10-20 ppb/day and that the error in this rate is at least 10% (due to ethylene concentration sampling error and leakage rate estimation error) and that the microbial rate is probably an extreme upper limit (especially since the BSAB uses plant waste as a substrate) it is safe to neglect BSAB ethylene in the total material balance.

VIII. Discussion of Results

Of the five technologies considered for reducing ethylene in the BPC only permanganate has the potential to meet the design constraints. The other methods were excluded for a variety of reasons, especially the potential to create a significant upset in BPC conditions. On the other hand, they could be revisited as future research developments warrant.

The permanganate removal rate was estimated to be about a half a ppb of ethylene per hour for 200 grams of material. Two factors need to be considered in determining the amount of material for BPC use; the amount of ethylene to be removed and the removal time. In the potato experiment the level needed to be reduced from about 60 ppb to about 30 ppb; however, for other crops such as wheat , this could be much higher.

A conservative estimate might be the removal of 100 ppb during a 10 hour period, i.e., a removal rate of 10 ppb/hour. For this rate the capacity of the removal system should be 4 Kg based on a constant ratio of mass of permanganate material to removal rate (the constancy of this ratio needs to be confirmed experimentally). Since the 4 Kg should be adequate for a very large number of removal episodes if kept unexposed to BPC air between episodes, the final system design should have the following features:

1) can be easily inserted into the duct for long term use
2) permanganate remains unexposed when not needed and readily exposed when ethylene removal is required
3) can be easily retracted from the duct for inspection (e.g., color change)
4) packing characteristics are time independent and resistant to high duct flow rates

IX. Conclusions and Recommendations

a) The current method for removing ethylene from the BPC during potato production by potassium permanganate absorption has a removal rate of approximately 0.5 ppb per hour.
b) Catalytic oxidation, molecular sieve, cryotrapping, and UV degradation were deemed unsuitable as alternatives to the current method.
c) A better way of removing ethylene with permanganate in the BPC might be the proper placement of 4 Kg of material in the air duct surrounding the BPC.
d) The Breadboard Scale Aerobic Bioreactor is not expected to contribute significantly, if at all, to the total material balance for ethylene in the BPC.
e) It is important to establish whether plants, during their normal growth period, respond to exogenous ethylene concentration.
f) Dosing and/or leakage experiments in the BPC might resolve this question.
g) If plants do respond to ethylene levels in the air space perhaps ethylene can be controlled through pulsing and removal to optimize the production process.
h) The current method for measuring ethylene concentration using a portable gas chromatograph exhibits variability which may reflect a diurnal cycle or could be due to sampling error.
i) In this regard it would be advantageous to study BPC air dynamics in more detail to determine the validity of the perfect mixing assumption.
j) A dry ice (carbon dioxide) drawdown experiment might be a good test of the perfect mixing assumption in the BPC.
k) Laser spectroscopy may afford an improved method for the measurement of ethylene concentration and should be explored for BPC use.
l) The dynamics of ethylene removal by permanganate need to be studied experimentally by simulating BPC conditions in a bench scale reactor. Mass transfer/reaction rate data for permanganate materials in a tubular system could be determined for different concentrations of ethylene which could lead to the establishment of a scaleup model.
m) The BPC (air duct or bypass) is a good test bed for evaluating ethylene removal technologies during periods of closure when the leakage rate is 10%.

X. References

1) Abeles, F.B., P.W. Morgan, and M.E. Saltveit, Jr. Ethylene in Plant Biology. 2nd Edition. Academic Press. 1992.

2) Eastwell K.C., P.K. Bassi, and M.E. Spencer. Comparison and Evaluation of Methods for the Removal of Ethylene and Other Hydrocarbons from Air for Biological Studies. Plant Physiology. Vol. 62. 1978. pp.723-726.

3) Jiang Y., L.V. Yentekakis, and C.G. Vayenes. Methane to Ethylene with 85% Yield in a Gas Recycle Electrocatalytic Reactor-Separator. Science. Vol. 264, June 10, 1994. p1563.

4) Menaucourt J. Pression De Vapeur Saturante De L' Ethylene Entre 77 K et 119 K. Journal de Chimie Physique. Vol. 79, #6, 1982. pp531-535.

5) Peterson B. Ethylene Data from Shuttle. Unpublished Data. Feb. 27,1995.

6) Shorter A.J., and K.J. Scott. Removal of Ethylene from Air and Low Oxygen Atmospheres with UV Radiation. Lebensm. Wiss. U. Technol. Vol. 19. 1986. pp 176-179.

7) Scott K.J. and R.B.H. Wills. Removal by UV Lamp of Ethylene and other Hydrocarbons Produced by Bananas. J. Sci. Food. Agr. Vol 22. 1971. pp 496-497.

8) Scott K.J. and R.B.H. Wills. Atmospheric Pollutants Destroyed in a UV Scrubber. Lab Practice. Vol. 22. 1973. pp 103-106.

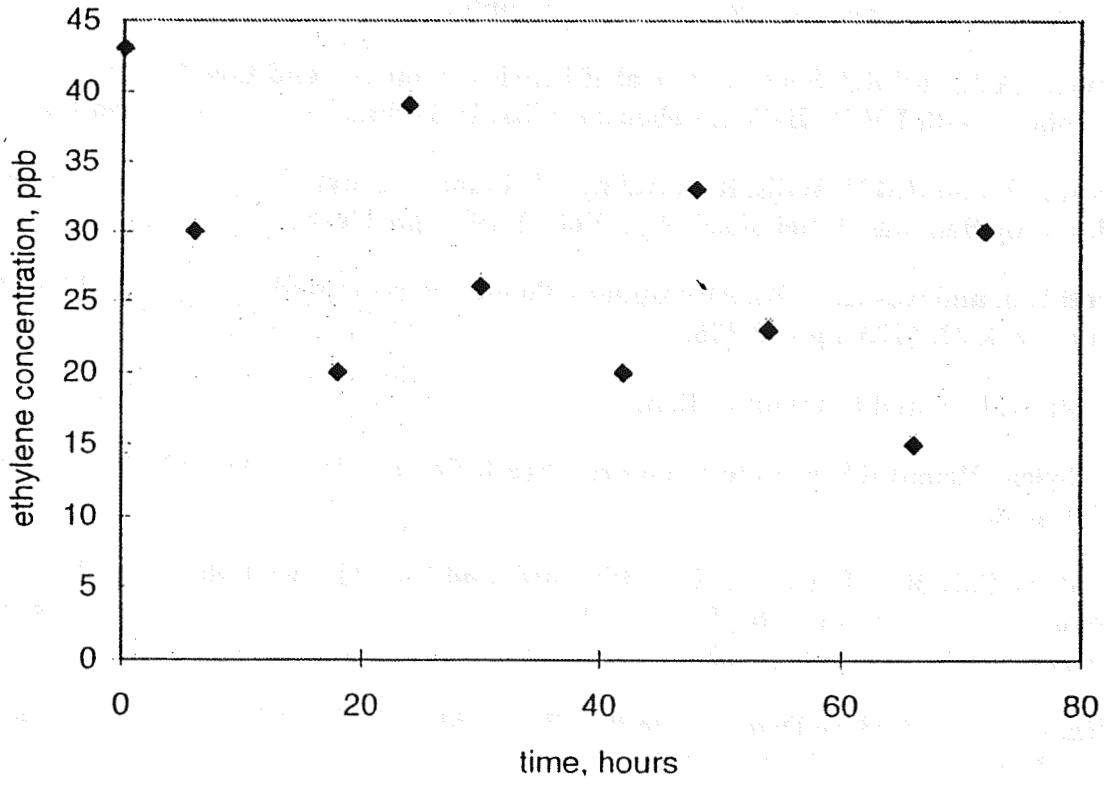
9) Knott W. Personal Communication.

10) Ethylene Removal System to Keep Fruit Fresh Twice as Long. Asian Defence journal. 10/1/91. p 86.

11) Lidster P.D., R.A. Lawrence, G.D. Blanpied, and K.B. McRae. Laboratory Evaluation of Potassium Permanganate for Ethylene Removal from CA Apple Storages. Transactions of the ASAE. Vol. 28. 1985. pp 331-335.

12) Rakow A.L. Ethylene Dynamics in the CELSS BPC. NASA/ASEE Summer Faculty Report. Kennedy Space Center. 1994.

Figure 1. Concentration vs. Time for Ethylene Removal by Permanganate



1995 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

S17-61
7757
P. 28
(28)

INVESTIGATION OF REGISTRATION ALGORITHMS FOR THE AUTOMATIC TILE PROCESSING SYSTEM

Dr. Dan E. Tamir
Associate Professor
Computer Science Department
Florida Institute of Technology
Melbourne, Florida

KSC Colleagues - Peter Engrand and Todd Graham
Artificial Intelligence/Expert Systems

Contract Number NASA-NGT-60002
Supplement 19

August 4, 1995

Acknowledgments

I would like to thank Peter Engrand, Todd Graham, and Terry Ross from NASA, as well as, Charlie Goodrich, Rich Bennett, and Bob Merchant from I-Net, for their cooperation and technical support. I would like also to extend compliments and thanks to Gregg Buckingham, Dr. Hosler, and Kari Stiles for their professional management of the summer faculty program.

Abstract

The Robotic Tile Inspection System (RTPS), under development in NASA-KSC, is expected to automate the processes of post-flight re-water-proofing and the process of inspection of the Shuttle heat absorbing tiles. An important task of the robot vision sub-system is to register the "real-world" coordinates with the coordinates of the robot model of the Shuttle tiles. The model coordinates relate to a tile data-base and pre-flight tile-images. In the registration process, current (post-flight) images are aligned with pre-flight images to detect the rotation and translation displacement required for the coordinate systems rectification.

The research activities performed this summer included study and evaluation of the registration algorithm that is currently implemented by the RTPS, as well as, investigation of the utility of other registration algorithms.

It has been found that the current algorithm is not robust enough. This algorithm has a success rate of less then 80% and is, therefore, not suitable for complying with the requirements of the RTPS.

Modifications to the current algorithm has been developed and tested. These modifications can improve the performance of the registration algorithm in a significant way. However, this improvement is not sufficient to satisfy system requirements.

A new algorithm for registration has been developed and tested. This algorithm presented very high degree of robustness with success rate of 96%.

Summary

This report summarizes the findings of the summer research performed at NASA-KSC. The research has concentrated on the registration algorithms for the Robotic Tile Processing System (RTPS).

First the image registration algorithm that is currently implemented by the RTPS is described and evaluated. Second, modifications to the existing algorithm are outlined. Next, a new approach is proposed. Results of experiments to evaluate the utility of the registration algorithms are presented and evaluated. Finally, conclusions that are based on the results of the experiments are drawn.

The registration procedure is divided into three phases; The first, referred to as center-to-tile is responsible for placing the camera so that one tile is completely included in an image frame. The second phase, performs coarse registration. The last phase, is responsible for fine registration.

The current system does not include a utility for tile centering.

The algorithm for coarse registration implemented by the current system is relatively simple. It applies one dimensional projections to an edge-enhanced tile-image. These projections are used for the identification of the tile vertices.

The fine registration phase relies on identification of tile "blobs"[1].

On-line and off-line experiments with the current algorithm has raised doubts about its adequacy for complying with the system requirements. On one hand, in many cases the algorithm has failed to identify the tile vertices. On the other hand, there have been concerns that the "blobs" can change significantly during the landing process, thus, confusing the fine registration procedure.

Modifications to the current algorithm has concentrated on the coarse registration process. Two modifications has been considered; First, it has been proposed to binarize the edge-enhanced image before obtaining the projections. If done properly, the binarization process can further emphasizes the tile edges and increase the rate of successful edge identification by the projections. The second modification includes the division of the tile into zones obtaining per-zone projections of the edge-enhanced image. The results of edge identifications per individual zone are then interpolated on the entire image to obtain a better approximation of the tile edges. These modifications have improved significantly the performance of the registration algorithm.

The modifications described above has prompted a probe of a new registration procedure. The idea has been to try to interpolate the edge lines directly from the binarized edge-enhanced image rather then using the zone projections for the interpolation. This idea has lead to the implementation of the Hough transform as an instrumental phase of the entire registration procedure.

In the new registration procedure, Hough transform is used to detect one edge, as well as, the circle that marks the re-water-proofing hole. This can be used for tile centering. Next, Hough transform is applied to the binarized, edge-enhanced, tile-image to identify the tile edges and vertices. Finally, cross-correlation of a small area around the tile vertices taken from the post-flight and pre-flight images is used for fine registration.

Experiments with the new coarse registration algorithm has yielded more than 96% success rate with an error of less than 0.4 degrees in the detected rotation and less than 0.1" in translation.

1. A "blob" is an homogeneous, continuous, region in the image.

Table of Contents

List of Figures

List of Tables

Introduction

NASA-KSC is developing a Robotic Tile Processing System (RTPS) for automatic post-flight position calibration, re-water-proofing hole verification, and visual inspection of the Shuttle tiles for anomaly detection.

The re-water-proofing and the inspection processes involve acquisition of pre-flight and post-flight images of the tiles, registration of these images, identification of the tile location for re-water-proofing, and detection of significant changes in the tile surface as a part of the tile inspection.

Image registration is one of the important tasks of the vision sub-system of the RTPS. Currently, the RTPS is applying two different registration algorithms. The first is used for the re-water-proofing hole verification. The second is a part of the automatic tile inspection.

A prototype of RTPS delivered to NASA-KSC has been tested and found deficient in its performance with respect to the image registration for re-water-proofing hole verification.

This summer research has concentrated on investigation of image registration techniques. In specific, the focus has been on the registration of the tile images for the re-water-proofing hole verification.

1.1 Background

Each Shuttle contains between 20,000 to 30,000 heat-absorbing tiles. Most of the tiles are square tiles with dimensions of 6"x6". Some tiles, however, has different sizes or shapes.

The robot, developed by CMU, is guided through a laser system. A second laser system on the robot arm places the arm in a given distance from the Shuttle surface and aligns the arm with the surface (three degrees of freedom, z, yaw, and pitch).

The robot vision sub-system, developed by SRI, is expected to take care of the other three degrees of freedom (x,y, and, roll) through the image registration process.

The system has been designed under the assumption that the total error in the identification of the robot location is relatively small (up to 1"). The data acquisition system has been designed to cover an area of 8"x8" with a resolution of 128 pixels/inch. Hence, under these assumptions, one entire tile is within one camera frame (one image).

It is currently assumed that due to changes in the Shuttle parking location and due to accumulated errors in the robot location system, there might be a discrepancy of up to 2" between the physical and the computed location of the robot. This new assumption emphasizes the need for a utility to center the RTPS camera around examined tiles.

1.2 The SRI Registration Algorithm

In addition to the data acquisition unit, the vision sub-system includes a CPU board (Motorolla 68040), a hard drive, and an image-processing board (Data-Cube). The system runs under a "UNIX-like" real-time operating system (VXWORK). Development tools include a C-compiler, a debugger (VXGDB), and a monitor. The system is accessible as an internet node.

The Shuttle surface is logically divided into zones of about 100 tiles per zone. Each zone is expected to include 3- 6 registration tiles. The process of registering the system to any of these registration tiles can be divided into three main procedures: Center_to_Tile, Coarse_Registration,

and Fine-Registration. These procedures are described briefly in the following section. A more detailed description is included in Appendix B of this document.

Center_to_Tile

Center_to_Tile is supposed to take care of the case where due to errors in the robot location sub-system the camera is not centered at a tile so that one or more of the edges of the tile are out of view. This module have not been fully developed by SRI. A module by this name exists, however, it fails if one or more of the tile edges is missing.

Coarse_Registration

This module attempts to identify the tile vertices through a process of edge enhancement and one dimensional projections. A tile vertex is defined to be the point where two tile edges coincide.

Edge enhancement is accomplished through convolution of the tile-image with an horizontal and a vertical gradient operator mask. Since the convolution requires extensive CPU time, this module has been ported by SRI to the Data-Cube. This is, however, the only computational intensive module ported to the Data-Cube.

After identifying the tile edges, the next modules of the registration software are applied only to the area enclosed within the edges. This has the effect of coarse registration.

Fine_Registration

Sometimes, it is difficult to determine whether a detected edge belongs to the tile in hand, or to the adjacent tile. Hence, a module for fine registration has been developed by SRI. This module attempts to identify continuous homogeneous regions in the tile (referred to as "blobs") and use these blobs for fine registration.

Segmentation is achieved through binarization of the images using a dynamic threshold. A connected component labeling algorithm identifies blobs. The blobs are classified according to their area, shape, and proximity to other blobs. Only blobs that comply with size shape and proximity restrictions are retained as "good" blobs. The centers of these blobs are used for the fine registration stage.

1.3 Modifications to the SRI Registration Algorithm

The modifications to the current algorithm has concentrated on the coarse registration process. The first modification is concerned with the edge detection part and the second attempts to on improve the procedure of obtaining the projections.

Edge Enhancement

Generally, the pixels that belong to the tile edges are darker than the rest of the image pixels. As a result, in many cases it is possible to skip the edge-enhancement phase and search for local minimums rather then local maximums in the projections. The advantage of this approach is that it reduces the computation time. For this reason, it has been investigated by KSC stuff. On the other hand, this method is the least robust of all the methods investigated so far. Moreover, given the available high-performance hardware, the reduction in computational complexity is irrelevant

Nevertheless, it is implemented (referred to as "ksc") as a part of the Off-line system described later in this paper.

Another modification relates to the edge enhancement procedure. In this version of the registration process, after the convolution the image is binarized using K-means clustering [3]. This approach has proved to be very robust with respect to identifying the edge vertices, however, it is inferior to the method described in the next section.

Per-zone Projections

The second modification to the SRI algorithm includes the division of the tile into zones obtaining per-zone projections of the edge-enhanced image. The results of edge identifications per individual zone are then interpolated on the entire image to obtain a better approximation of the tile edges. These modifications have improved significantly the performance of the registration algorithm. Moreover, this modification has prompted a probe of a new registration procedure. The idea has been to try to interpolate the edge-lines directly from the binarized edge-enhanced image rather than using the zone projections for the interpolation. This idea has lead to the implementation of the new algorithm for registration described in the next section.

A New Registration Algorithm

In the new registration procedure, Hough transform is used to detect one edge, as well as, the circle that marks the re-water-proofing hole. This is used for tile centering. Next, Hough transform is applied on the binarized, edge-enhanced, tile image to identify the tile edges and vertices. Finally, cross-correlation of a small area around the tile vertices of the post-flight and pre-flight images is used for fine registration. The new algorithm is described briefly in the following sections. A detailed description is available in Appendix B.

2.1 Center-To-Tile

A module for centering the camera around one tile has been developed. Two versions of the module exist. The first is using a circular Hough transform to identify the circle marking the re-water-proofing hole [1]. This module has been tested on a set of tile images that is currently available. It has been found to identify the circle in 63 out of 66 tile images.

A second version for tile centering procedure relies on the linear Hough transform [2]. It attempts to identify at least one vertex and use this vertex for centering the camera. This version has not been tested yet.

2.2 Coarse-Registration

The module Center_to_Tile, as well as, the module Coarse_Registration are using Hough transform. The transform is applied to a binary image obtained from edge enhancement and line enhancement kernels. The difference between these two modules is that Center_to_Tile is using edge enhancement and circular Hough transform, while Coarse_Registration is using line enhancement and linear Hough transform. These operation are described below.

In the future, experiments can be conducted to examine the utility of applying other convolution masks including noise reduction operators.

After the convolution the image is binarized using K-means clustering [3]. The algorithm is executed with K=3, hence, it partitions the edge-enhanced/line-enhanced pixels into three classes. It has been found that class three of the edge-enhanced pixels are best for the circle identification while class two of the line-enhanced pixels are best for the tile edges detection.

The Hough transform is considering every pixel in the appropriate cluster. The circular transform attempts to draw a circle around every pixel from class three of the edge enhancement image. It measures how many of these circles coincide. The linear Hough transform is attempting to draw as many lines as possible through each pixel in class two of the line-enhanced image. It measure how many of these lines coincide. It is assumed that most of the circles with the same radius as the radius of the circle that marks the re-water-proofing hole coincide at the center of this circle. The lines going through tile edges share the same slope and intercept. Hence, most of the lines coincide in slopes and intercept that correspond to the slope and intercept of the tile edges. This can enable the detection of the tile vertices.

The linear Hough transform has been applied to 156 tile images and has identified the lines in 96.5% of these images with an accuracy of about 0.1" in detecting translation and 0.4 degrees in detecting rotation.

2.3 Fine-Registration and AMDF

One concern with the fine registration method developed by SRI is that the blobs, used as a part of the fine registration, are susceptible to changes between flights.This is due to the effect of the heating of the tiles when the Shuttle returns to the atmosphere and the collision of gases and dust with the tiles during the landing. One assumption has been that the tile vertices are less sensitive to these problems, hence, they can be used for a more robust fine registration. This assumption has lead to a modified fine registration algorithm.

The Average Magnitude Difference Function can be described as a brute-force approach for registration [4]. This method ranks every possible rotation and translation and selects the best. Being a brute-force method, it can not be considered for implementation on large images, however, after the detection of the tile vertices, AMDF can be applied to a set of small images centered around the vertices. This procedure has been implemented, but, has not been fully tested yet.

Experiments, Results, and Evaluation

An off-line system for experiments with the different algorithms has been developed. The Off-line system is a set of C procedures. It is further described in appendixes A and B.

The experiments, the results, and result evaluation are given below.

3.1 Experiments

Two sets of images have been acquired using the SRI camera and acquisition software. The first set consists of 66 images of 22 tiles mainly from one zone of the Shuttle Endeavor. The second set includes 90 images of 30 tiles acquired from the Shuttle Columbia. This set includes "tough" tiles from different zones of the Shuttle. They are considered "tough" since according to our subjective judgement each of them contains several artifacts that might impose difficulty in the process of identifying the tile-edges for one or more of the tested algorithms. Each tile has been acquired with three different lighting set-ups. In addition, a few more images of some of the tiles have been acquired with some degrees of rotation and translation.

All the algorithms excluding the per-zone method have been applied to the first set of tiles. Only the new algorithm has been applied to the second set. The results of these experiments are summarized in the next section.

3.2 Results

The results of the experiments are summarized in Table 1. The entries of the table list the percentage of correct identification of the tile edges and the average CPU time (seconds). In this table, "SRI" refers to the method implemented by SRI and "KSC" refers to a modification to the SRI method. Under this modification, edge-enhancement is not applied to the tile-image. Instead, the algorithm obtains the projections of the tile-image and searches for minimums. The method "Trin" refers to the case in which the image goes through edge-enhancement, and binarization using 3-means clustering before the one dimensional projections are obtained. The method "Line" refers to the application of the linear Hough transform, and the method "Circle" refers to the application of the circular Hough transform.

Set/Method	SRI	KSC	Trin	Line	Circle
First Set	81% / 190	77% / 10	100% / 250	100% / 310	95% /310
Second Set	n/a	n/a	n/a	93% / 310	n/a
Total	n/a	n/a	n/a	96.5% / 310	n/a

Table 2: Experiments and Results

3.3 Evaluation

The evaluation of these methods analyses two aspects of performance: identification accuracy and computational complexity. The identification accuracy is evaluated using the error rate of individual method and the accuracy of identifying the edge parameters. The computational complexity is evaluated based on the average CPU time obtained in these experiments.

Identification Accuracy

The experiments include three methods for identifying the vertices using one dimensional projections. Among this three, the method referred to as "Trin" is a clear winner.

One problem with these three methods is that they are sensitive to rotation. Moreover, in their current form, they do not give a good indication of the edge parameters. That is the distance of the tile vertices from the origin of the coordinate system[1] and the angles between each of the tile vertices and the X -axis of the coordinate system. Hence, they require an additional step.

The procedure "Line" does this additional step at the same time that it identifies the tile edges. This procedure identifies the edges and supplies their distance from the origin and angle with the X -axis of the model coordinates. The accuracy of "Line" is about 0.4 degrees and 0.1". The method "Line" has the same or better error rate compared to "Trin", this is done with additional, relatively accurate, information. For this reasons, the method "Line" is considered the best method tested so far.

The method "Circle" has identified the circle around the re-water-proofing hole in 95% of the times. There are, however, two concerns with respect to the performance of this method; First, it relies on an artifact (the circle) that can change significantly between flights. Second, it does not identify the center of the circle accurately enough (an error of about 0.25" is expected). This method, however, has been designed for the module center-to-tile and not for the coarse registration. Therefore, it should be compared with other methods for centering the tile-image. This is one of our recommendations for future research.

Computational Complexity

Based on the information in Table 1, the algorithms are ranked according to their complexity in the following ascending order (the lower the better):

1) "KSC"

2) "SRI"

3) "TRIN"

4) "Line" and "Circle"

These figures are based on running the methods off-line on a Sun Sparc -10. However, with respect to the RTPS, the issue of complexity should not be considered in the process of selecting an algorithm since the actual system contains a very powerful pipeline-processor, the Data Cube. Actually, the procedure "SRI" is already implemented on the Data-Cube and runs in less than 5 seconds. Moreover, the Data-Cube contains a module for Hough transform which can be used for implementing the procedure "line".

1. The coordinate system referred to here, is the coordinate system of the shuttle model stored in the robot.

Conclusions and Recommendations for Future R&D

4.1 Conclusions

Based on the evaluation of the results of the experiments., the following conclusions are drawn:

- Edge-enhancement is an essential operation. It should not be skipped (as in the "KSC" algorithm). The current edge-enhancement masks give reasonable results, however, there is still need for further research into other edge-detection and edge-enhancement methods.
- After edge-enhancement, the image has to be binarized in order to further emphasize the edge information. The binarization has to be performed in a method similar to the one investigated in this research.
- Among the methods investigated so far, the linear Hough transform is the most appropriate method for identifying tile-edges in the process of coarse registration.
- Identifying the circle using the circular Hough transform has some potential for the procedure center-to-tile. However, it is possible that using the linear Hough transform in an attempt to identify one tile vertex will have better utility for center-to-tile.

4.2 Recommendations for Future R&D

It is recommended that future research and design concerning the registration system will include the following components:

- Experiments to examine the utility of applying other convolution masks including noise reduction operators.
- Experiments to determine the utility of identifying at least one vertex using this vertex for centering the camera on a tile.
- Experiments to asses the utility of cross-correlation or AMDF in the process of fine registration [3].
- It is recommended to implement the new registration algorithm on the robot vision system. The development should include three main phases: testing, implementation on the Motorolla 68K board, and implementation using the Data-Cube.
    - The testing process can be done both off-line (on a Sparc work-station) and on-line (using the robot vision system). The main goals are to determine the best light configuration, and to finalize the details of the algorithms to be implemented on-line. It is proposed to obtain about 150 tile images from the Shuttle Columbia, as well as, about 150 tile images from the Shuttle Endeavor. Part of the Endeavor images will cover the same tiles obtained during this summer. These images will be used as "post-flight" images.
    - The second phase is to integrate the new algorithms with the current robot vision software on the 68K processor.
    - Finally, some of the modules of the new algorithm should be implemented on the Data-Cube. For example, the AMDF and the clustering algorithms are relatively time consuming. These algorithms can be ported to the Data-Cube.

Bibliography

[1] R. C. Gonzalez and P. Wintz, *Digital Image Processing,* (2'nd Ed.), Addison Wesley, Reading, MA, 1987.

[2] S. A. Dudani and A. L. Luk, "Detecting Straight Line Segments on Outdoor Scenes", *Pattern Recognition,* vol. 10, 1987: 145- 157.

[3] L.R., Rabiner and R. W. Schafer, *Digital Processing of Speech Signals,* Prentice Hall, Englewood Cliffs, NJ, 1978.

[4] J. T. Tou and R. C. Gonzalez, *Pattern Recognition Principles,* Addison Wesley, Reading, MA, 1974.

[5] Dan E. Tamir, Chi-Yeon Park and Book-Sung Too, "Vector Quantization and Clustering: A Pyramid approach", *IEEE Data Compression Conference,* Utah, March, 1995.

Appendix A: The Off-line Programs - A User Manual

Appendix A contains instructions for the user of the Off-line programs. These instructions include installing, compiling and running the Off-line programs.

A.1 Installing the Off-line programs

The source code of the Off-line programs and a "make-file" are stored in an archive in the Unix "tar- file" "off-073095.tar". To install the programs, place the file "off-073095.tar" in any directory and apply the command - "tar xvf off-073095.tar". This command extracts the following set of files into the directory:

1) Makefile
2) main.c
3) image_read.c
4) image_process.c
5) trinarize2.c
6) trinarize3.c
7) hconv.c
8) vconv.c
9) line_enhance.c
10) point_enhance.c
11) image_write.c
12) hproj.c
13) vproj.c
14) image_add.c
15) image_sub.c
16) circle_find.c
17) line_find.c
18) register.h

A.2 Compiling the Off-line program

The Off-line program consists of one user interactive program ("register") that is geared to enable the user to test several registration algorithms, and six individual programs each testing a specific registration algorithm. These programs are 1) "org" - a program to generate a sun-raster-file from a file acquired by the SRI camera, 2) "sri" - a program to test the coarse registration algorithm developed by SRI, 3) "ksc" - a modification to the SRI algorithm that has been developed by KSC stuff, 4) "trin" - a modification to the SRI algorithm that binarizes the edge-enhanced image, 5) "line" - a registration algorithm that is based on a linear Hough transform, and 6) "circle" - a frame centering algorithm that is based on a circular Hough transform.

The file "Makefile" is used to compile the Off-line programs. The command "make" or "make all" creates the entire set of programs ("register", "org", "sri", "ksc", "trin", "line", and "circle"). This command invokes the Sun "cc" compiler and linker on the source-code and generates object files and executable files.

To compile a specific program such as "register" or "sri" use the command "make register" or "make sri".

In addition, the "make-file" contains an instruction to create an archive of the source-code (C code, Makefile, and include files). To activate this command use the instruction "make tar". Another instruction available in the "make-file" is "make clean". This instruction removes all the object files and the executable files generated in the compilation stage.

A.3 Running the programs

The set of programs Off-line requires more than 8MB of stack space to run. Thus, before running the programs the user should check his stack space using the command "limit". If needed, the size of the stack space can be increased to 16384k Bytes using the command "limit stacksize 16384k".

Detailed description of the algorithms along with flow-charts of the programs are available in appendix B.

The program "register" is a menu driven program. It prompts the user for input files which are either Sun-rater-files or camera acquired files of the tiles. It produces an output file that contains the result of applying a specific algorithm to the input image. The output file is in Sun-raster format. The file name is given by the user. The following algorithms are implemented in "register":

1) An algorithm to create a Sun Raster File.

2) An algorithm for Line Enhancement.

3) An algorithm for Point Enhancement.

4) An algorithm to binarize an image using 3-means clustering and choosing cluster - 2.

5) An algorithm to binarize an image using 3-means clustering and choosing cluster - 3.

6) An algorithm to implement SRI method for detecting the tile edges.

7) An algorithm to implement KSC modification to the SRI method.

8) An algorithm to implement the binarization modification to the SRI method.

9) An algorithm to detect the circle around the re-water-proofing hole.

10) An algorithm to detect the tile edges using a linear Hough transform.

11) A procedure to add two images.

12) A procedure to produce an error image using the absolute difference between the images.

Some of these algorithms requires pre-processing by other algorithms in the set. For example, before applying the algorithm to detect the tile edges the user has to run the algorithm for line enhancement and the algorithm for binarization with cluster two. The menu instructs the user on the proper way of activating each of these routines.

In addition to the interactive program "register", a few programs that can be run from the Unix command line or from shell files are included in the Off-line programs. Unlike the interactive program "register", where some algorithms require pre-processing, the line-command programs are self-contained. That is, they operate directly on the input image and do not depend on the result of pre-processing of the input image. The following sections explain how to execute these programs.

The program "org" is executed using the command "org in out" where "in" is the name of an existing tile-image acquired from the SRI camera and "out" is a new name. The program converts the image stored in "in" into Sun raster format and stores it in the file "out".

The program "sri" is executed using the command "sri in out" where "in" is the name of an existing tile-image in Sun raster format and "out" is a new name. The program applies the SRI algorithm to the image "in" and presents the results in an image stored in Sun-raster format in the file "out".

The program "ksc" is executed using the command "ksc in out" where "in" is the name of an existing tile-image in Sun raster format and "out" is a new name. The program applies the KSC modification to the SRI algorithm to the image "in" and presents the results in an image stored in Sun-raster format in the file "out".

The program "trin" is executed using the command "trin in out" where "in" is the name of an existing tile-image in Sun raster format and "out" is a new name. The program applies edge-enhancement, binarization and one dimensional projections to the image "in" and presents the results in an image stored in Sun-raster format in the file "out".

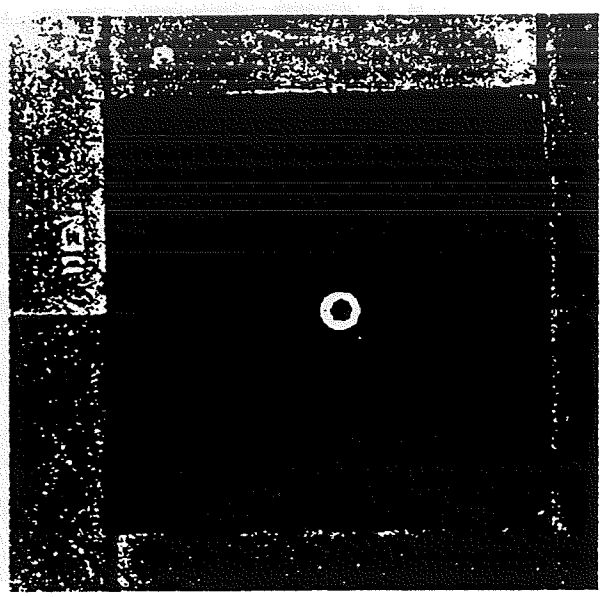
The program "circle" is executed using the command "circle in out" where "in" is the name of an existing tile-image in Sun raster format and "out" is a new name. The program applies a circular Hough transform to the image "in" and attempts to identify the circle around the re-waterproofing hole. The program "circle" presents the results in an image stored in Sun-raster format in the file "out".

The program "line" is executed using the command "line in out" where "in" is the name of an existing tile-image in Sun raster format and "out" is a new name. The program applies a linear Hough transform to the image "in" and attempts to identify the tile edges. The program "line" presents the results in an image stored in Sun-raster format in the file "out".

The output of the Off-line modules is an image in Sun raster-file format. Image display is accomplished through the shareware XV.

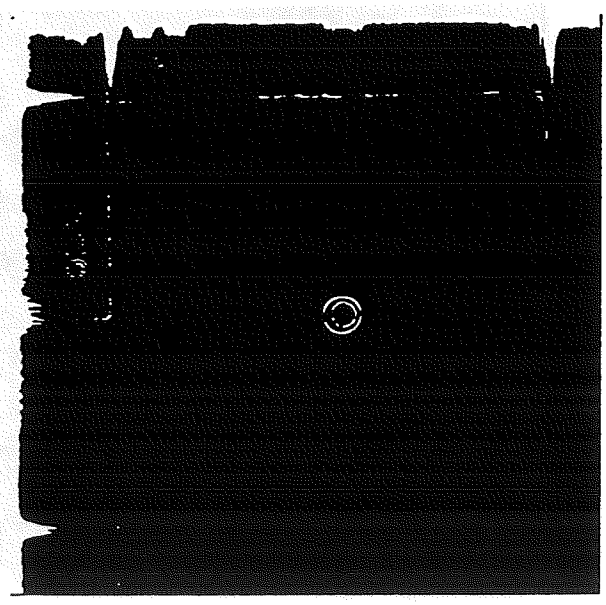
The following Figures illustrate the results of applying these algorithms to a tile-image. Figure 1a, contains the original image. Figure 1b, contains the results of applying edge-enhancement and projections (SRI method) on the image of Figure 1.a. Figure 1.c, is the result of superimposing the Figures 1.a and 1.b. Finally, Figure 1.d. is the result of applying the program "ksc" on the image of 1.a.

Figure 2, contains the line-enhanced image along with the results of running "trin" "circle" and "line".

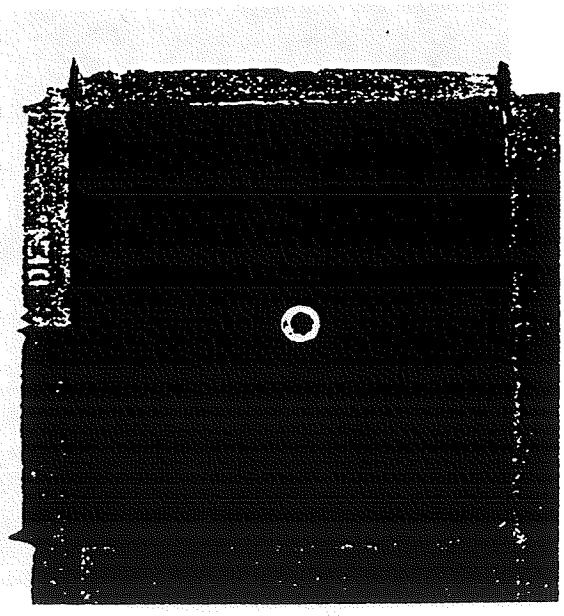


a) The Tile-Image

b) Result of Convolution and Projections

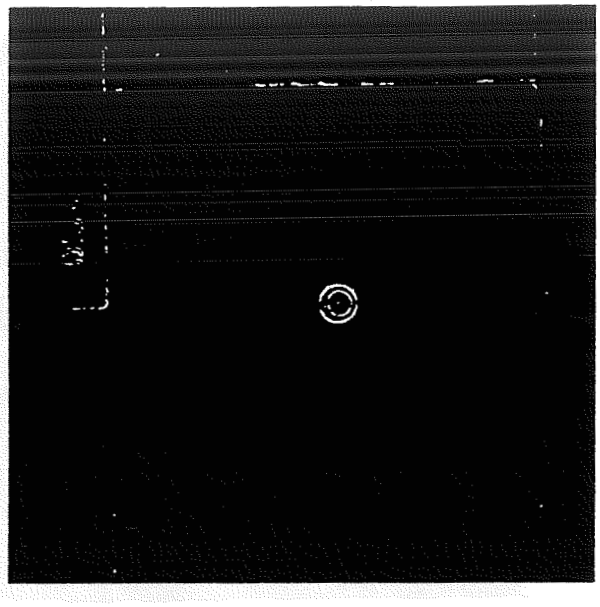


c) Output of "sri"

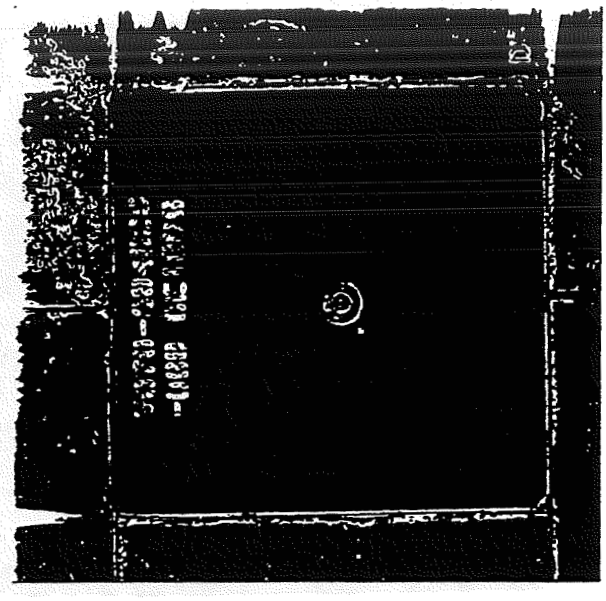


d) Output of "ksc"

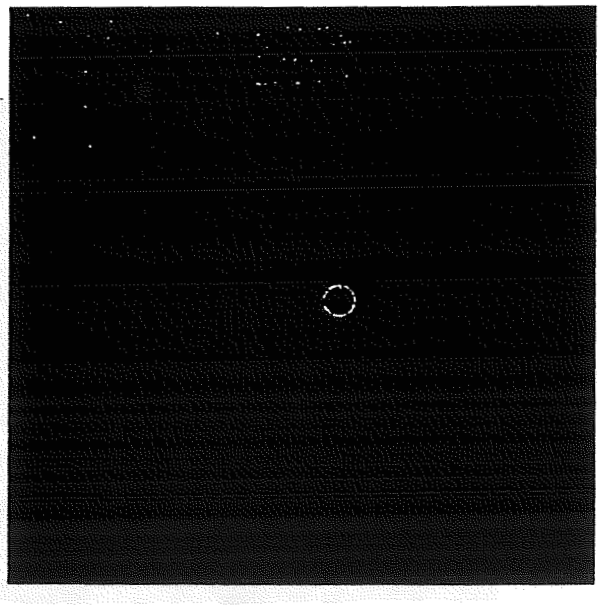
FIGURE 1. Output of "ksc" and "sri"



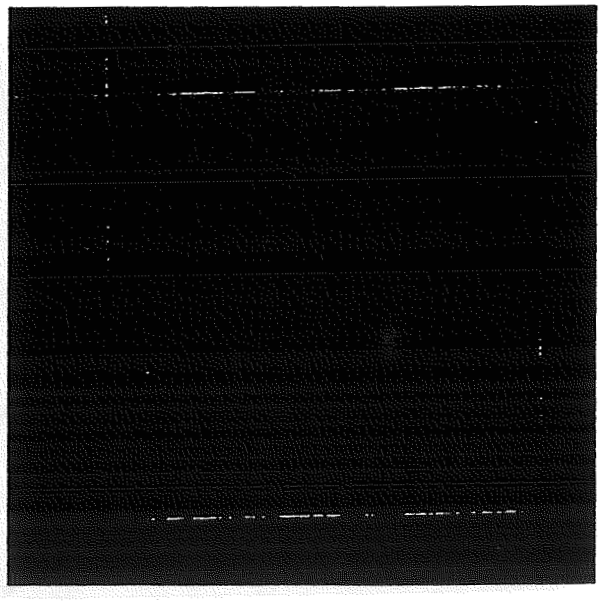
a) Output of "line_enhance"



b) Output of "trin"



c) Output of "circle"



d) Output of "line"

FIGURE 2. Output of "line_enhance", "trin", "circle", and "line"

Appendix B: The Off-line Programs - A Programmer Manual

Appendix B contains instructions for a programmer, with basic knowledge in image processing and C, who has to modify or maintain the Off-line system. The manual explains the function of the important modules included in the Off-line programs.

B.1 The program main.c

This program checks the command line arguments. If the first argument is "register" then the program enters into a switch loop where a menu is printed and selection of different items from the menu activates the program module image_process.c with an indication of the menu item selected. Image process activates specific image processing modules according to the flow diagrams of Figures 3, and 4.

If the command is "org", "sri", "ksc", "trin", "circle", or "line", then the program checks that the number of arguments is three and activates the appropriate algorithm according to Figures 3 and 4. Any other activation of the main program terminates the program.

B.2 The program image_process.c

The module image_process.c has three parameters, an integer which denotes the algorithm to be activated an input matrix that is filled by image_read.c and an output matrix that is filled by the activated algorithm. A switch statement selects the algorithm to be activated according to the parameter transferred from main.c.

B.3 The program image_read.c

The input to image_read.c is a string that contains the name of an existing file. The output is a "1024 by 1024" matrix of unsigned characters. This program opens the input file (no error checking) and checks the first 4 Bytes. If these Bytes corresponds to the Sun-raster "magic" number then the program skip the next 766 characters which are assumed to belong to the header and color-map of the file. The program then reads the next "1024 by 1024" pixels into the output array. If the first four Bytes does not match "magic", then it is assumed that it is a file in "SRI" format. That is, an image file obtained and saved through the SRI system. In this case, the program skips the next 1020 Bytes and reads "1024 by 1024" pixels into the output matrix.

B.4 The program image_write.c

The input to image_wirte.,c is a string that represents a name of a file to be created and a "1024 by 1024" matrix of unsigned characters. The program creates a file in Sun raster format (no error checking). The file contains "1024 by 1024" one-Byte pixels. Before writing the image pixels, an appropriate Sun raster header and a color-map that contains identity mapping are written into the file. By identity mapping it is meant that the (R,G,B) vector in location "i" of the color map contains the values (i, i, i). Hence a pixel with the value "i" is mapped into the gray level "i".

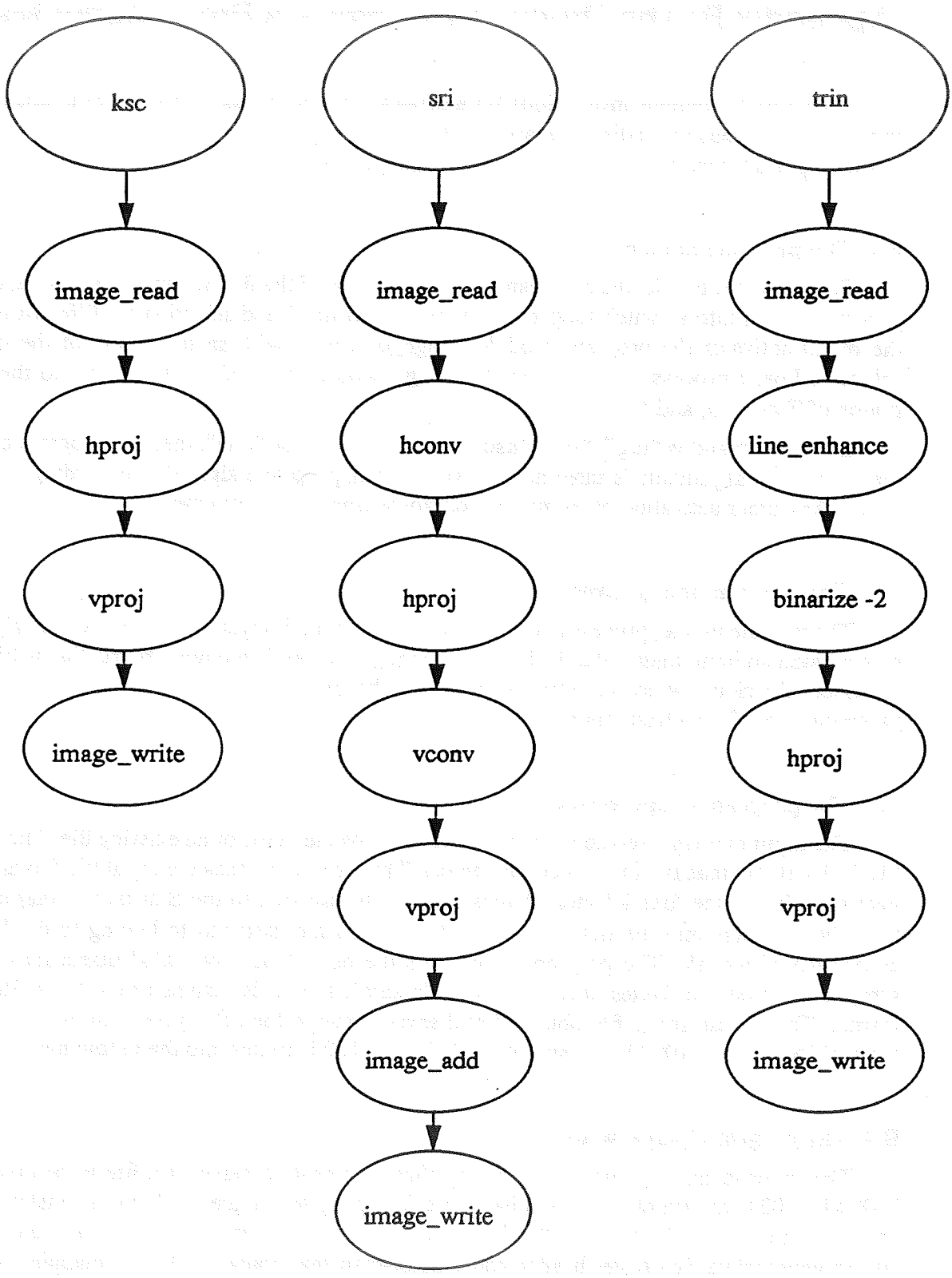


FIGURE 3. The nodule "ksc", "sri", and "trin".

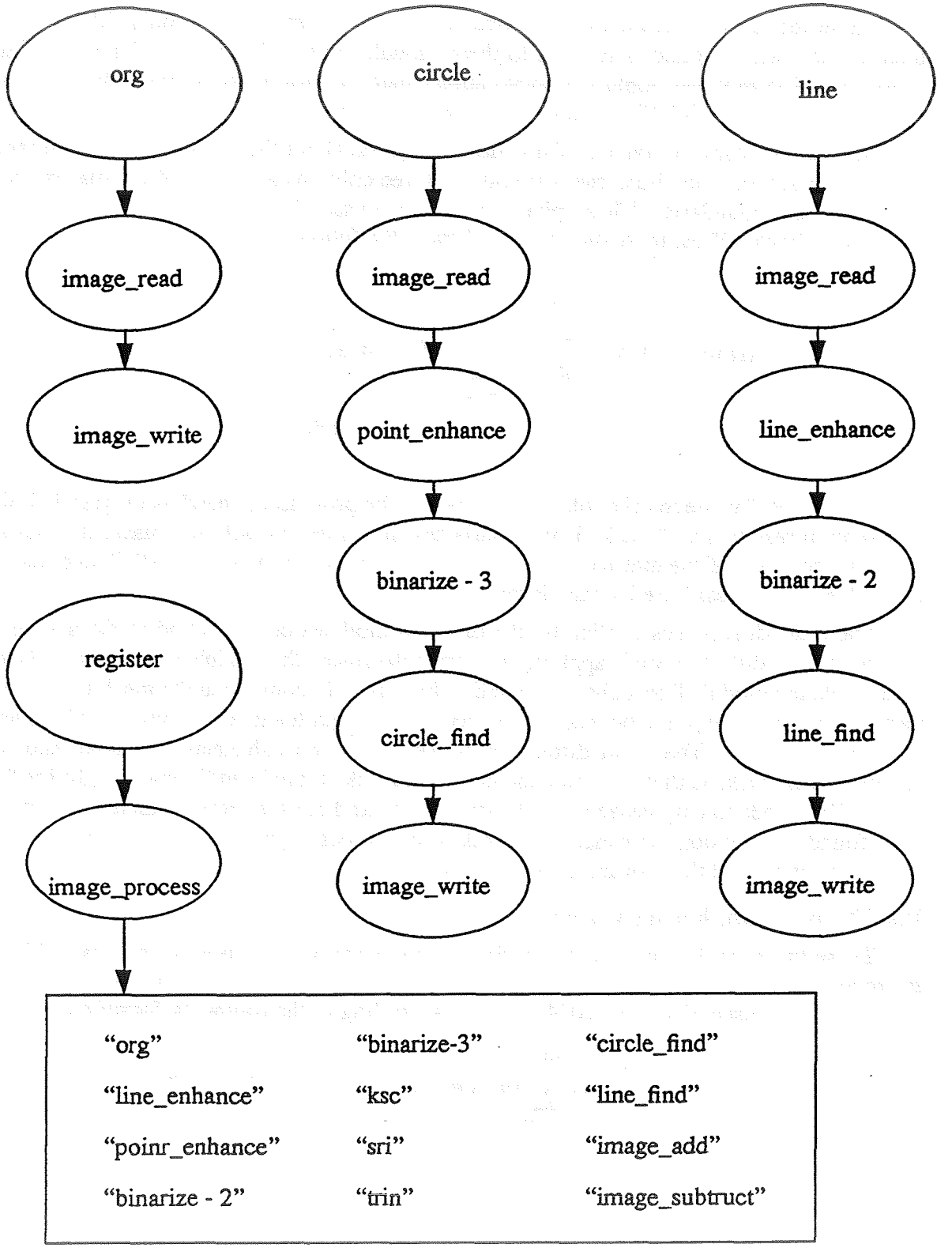


FIGURE 4. The Modules: "org", "circle", "line", and "register"

B.5 The programs hconv.c, vconv.c, line_enhance.c, and point_enhance.c

These modules implement convolution with "7 by 7" masks. The only difference between them is the masks applied. The input to these modules is a "1024 by 1024" matrix of unsigned characters. The modules apply the convolution, scale the results to the range [0,255], and write them into a "1024 by 1024" matrix of unsigned characters.

As with any convolution, a decision has to be made about the boundaries of the image. In the implementation reported here, the first and last three columns and rows of the matrix are considered to be the boundaries. These pixels are not processed by the module, effectively, they are replaced by zeros. Thus, these modules implement the following formula:

$$new(i,j) = \sum_{k=-3}^{3} \sum_{l=-3}^{3} old(i+k, j+l) \times mask(k,l)$$
$$(i,j) = 3, 4, \ldots, 1020$$

Where "old" is the matrix holding the image to be processed, "new" is the matrix holding the convolution results, and "mask" is the matrix that holds the convolution mask. It is assumed the rows and columns of the matrices "old" and "new" are indexed from 0 to 1023, and that the rows and columns of "mask" are indexed from -3 to 3.

The individual masks applied by the different modules can be found in the source code. In general, the module hconv.c is applying an horizontal mask, the module vconv.c is applying a vertical mask, the module line_enhance.c applies the mask of hconv.c and the mask of vconv.c in one path, and the module point_enhance.c is applying in one path a modified version of the horizontal and vertical masks. The main difference between the line enhancement masks and the point enhancement masks is that the point enhancement masks include additional weight for the center pixels. This tends to emphasize a small local change and not just horizontal/vertical edges. It has been found that the point enhancement masks produce better pre-processing for the identification of the circles around the re-water-proofing holes.

B.6 The programs hproj.c vproj.c

These two modules compute the horizontal and vertical projections of an image. The input to the modules is a "1024 by 1024" matrix of unsigned characters. The programs generates an array (hproj or vproj respectively) of 1024 elements according to the following formulae:

$$\left(hproj(i) = \sum_{j=0}^{1023} image(j,i) \right) \; ; i = 0, 1, 2, \ldots, 1023$$

$$\left(vproj(i) = \sum_{j=0}^{1023} image(i,j) \right) \; ; i = 0, 1, 2, \ldots, 1023$$

Where "image" is a "1024 by 1024" matrix, "hproj", and "vproj" are arrays of rank "1024". The modules hconv.c and vconv.c scale the arrays "hconv" and "vconv" to the range [0,127] and

then generate an output matrix in which the first 128 rows/columns contain an "histogram" representation of the results of the projections. Figure 1.b from page 19 illustrates this histogram.

B.7 The programs trinarize2.c and trinarize3.c

These programs implement the K-means algorithm with the LBG termination condition on the one dimensional data set obtained form the gray-level values of the pixels in an input image [4], [5]. In this application "k" is selected to be 3 and the initial cluster centers are 0, 127, and 255, respectively. It is assumed that after convergence, the first cluster center represents the background, the second cluster center correspond to weak edges, and the third cluster center corresponds to strong edges. This assumption has been confirmed through a large set of experiments.

Strong edges correspond mainly to the white characters, white circles, and other white marks, drawn on the tile by the Shuttle maintenance crew. Therefore, applying point enhancement and 3-means clustering to the image and then picking cluster 3, filters out most of the information from the image and leaves just the characters, circles, and other white marks. This is exactly what trinarize3.c is doing. It is generating a binary output image in which every pixel that has been clustered into cluster 3 is assigned the value 255. Other pixels are assigned the value 0. This image can be used for the circle identification (see Figure 2.c).

Weak edges correspond mainly to the edges of the tile. Therefore, applying line enhancement and 3-means clustering on the image and then picking cluster 2, filters out most of the information from the image and leaves just the tile edges. This is exactly what trinarize2.c is doing. It is generating a binary output image in which every pixel that has been clustered into cluster 2 is assigned the value 255. Other pixels are assigned the value 0. This image can be used for the line identification (Figure 2.d).

The current version of trinarize2.c and trinarize3.c have not been optimized for performance. If these modules are to be used in the on-line system, it is recommended to either use the Data-Cube hardware for the implementation, or to use a fast version of K-means developed by Tamir [5].

B.8 The program circle_find.c

This program operates on a binary image. It is modeling the operation of "drawing" a circle around every non-zero pixel in the input image and uses the results of this modeling to update a scratch-pad matrix. Let (y,x) be the coordinates of a non-zero pixel in the original image. The coordinates (y_i,x_i) of the points that lie on a circle with radius "R" around (y,x) are computed using the following formula:

$$(y_i = y + \lfloor R \times \sin(i) \rfloor)\ (x_i = x + \lfloor R \times \cos(i) \rfloor)\ i = 0, 1, 2, \ldots, 180$$

In this formula, $\lfloor x \rfloor$ represents the floor of "x". The radius of most of the circles around the re-water-proofing holes is approximately 28 pixels. Hence, "R" is set to 28 in the module circle_find.c.

The scratch-pad matrix "accum" is a "1024 by 1024" short integers matrix. It is initialized to contain zeros in all the places. The module circle_find.c computes the coordinates of all the points (y_i,x_i) that lie on circle with a radius of 28 pixels around every non-zero pixel in the input image. For every computed point (y_i,x_i) the module circle_find.c increments the entry (y_i,x_i) in the matrix "accum" by one. In the next stage, the function circle_find.c searches for the maximum value in

"accum" assuming that the location with maximum value corresponds to the center of the re-water-proofing hole verification circle. The program then dims the original image by dividing every gray level by two and draws a white (gray-level of 255) circle with a radius of 28 pixels in the original image around the identified center. The resultant image is the output of the program. Figure 2.c from page 19 illustrates the output of circle_find.c with respect to the image from Figure 1.a.

B.9 The program line_find.c

This program operates on a binary image. It is modeling the operation of "drawing" lines that pass through non-zero pixel in the input image and uses the results of this modeling to update a scratch-pad matrix. In this application, a line is represented using the normal representation [1]. Under this representation the (y,x) coordinates of the points that lie on the line correspond to the following equation:

$$x \times \cos(\theta) + y \times \sin(\theta) = \rho$$

Where θ is the angle between the x-axis and a vector perpendicular to the line, and ρ is the distance of the line from the origin.

Let (y,x) be the coordinates of a non-zero pixel in the original image. The program line_find.c evaluates several lines that pass through (y,x) these lines correspond to the equation:

$$x \times \cos(\theta_i) + y \times \sin(\theta_i) = \rho_i$$

Under the current assumptions about the accuracy of the robot location system, the possible range for ρ in pixels is [0,255] and [768,1023]. The possible range for θ in degrees is [-5°,5°] and [85°,95°]. However, to increase accuracy, the angle θ is measured with an accuracy of 0.2 degrees. Thus, given a non-zero pixel in location (y,x), the program line_find.c computes the integer values of all the ρ_i and θ_i, such that the point (ρ_i,θ_i) complies with the accuracy constraints and satisfies the equation $x \times \cos(\theta_i) + y \times \sin(\theta_i) = \rho_i$.

The values of ρ_i in the range [768,1023] are mapped to the range [256,511]. The values of θ_i in the range [-5,5] (with accuracy of 0.2) are mapped into the range [0,49] and the values of θ_i in the range [85,95] (again, with accuracy of 0.2) are mapped into the range [50,99].

The scratch-pad matrix "accum" is a "512 by 100" matrix of short integers. It is initialized to contain zeros in all the places. For each non-zero pixel at location (y,x) the module line_find.c computes the (ρ_i,θ_i) parameters of all the lines that pass through (y,x) and comply with the distance and angle constraint. Let (ρ_i,θ_i) be such a pair of parameters. The module line_find.c increments the corresponding entry in the matrix "accum" by one. In the next stage, the function line_find.c searches for maximum values in four ranges of "accum" assuming that the locations with maximum values corresponds to the individual edges of the tile. The ranges are:

1) $((0 \le \rho_i \le 255), (0 \le \theta_i \le 49))$,
2) $((256 \le \rho_i \le 511), (0 \le \theta_i \le 49))$,
3) $((0 \le \rho_i \le 255), (50 \le \theta_i \le 99))$, and
4) $((256 \le \rho_i \le 511), (50 \le \theta_i \le 99))$

The program then dims the original image by dividing every gray level by two and draws inside the dimmed image four white (gray-level of 255) lines corresponding to the (ρ_i,θ_i) of the four maximum values. The resultant image is the output of the program. Figure 2.d from page 20, illustrates the output of line_find.c with respect to the image from Figure 1.a.

B.10 The program image_add.c

The input to this program is two "1024 by 1024" matrices of unsigned characters. It is applying the following formula:

$$new(i, j) = max(255, (old1(i, j) + old2(i, j)))$$

where "old1" and "old2" are the two input matrices, "max" stands for maximum, and "new" is the output matrix.

B.11 The program image_sub.c

The input to this program is two "1024 by 1024" matrices of unsigned characters. It is applying the following formula:

$$new(i, j) = abs(old1(i, j) - old2(i, j))$$

where "old1" and "old2" are the two input matrices, "abs" stands for absolute value, and "new" is the output matrix.

B.12 The header file register.h

This file contains the following definitions:

```
# include <stdio.h>
# include <math.h>
# include <memory.h>
# include <string.h>
# include <pixrect/pixrect_hs.h>
# define PI 4 * atan(1.)
# define RAS_MAGIC 0x59a66a95
# define MAGIC 0xa659956a
# define DEPTH 8
# define TYPE 1
# define MAPTYPE 1
# define MLENGTH 768
#define IMAGE_X 1024
#define IMAGE_Y 1024
```

1995 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

S/B-81
7758
I-36
36

PRIORITIZATION OF ENGINEERING SUPORT REQUESTS AND ADVANCED TECHNOLOGY PROJECTS USING DECISION SUPPORT AND INDUSTRIAL ENGINEERING MODELS

Dr. Madjid Tavana
Assistant Professor and Chair
Management Department
La Salle University
Philadelphia, Pennsylvania

KSC Colleague - Seunghee Lee
Industrial/Business Management

Contract Number NASA-NGT-60002
Supplement 19

August 1, 1995

PAGE 500 INTENTIONALLY BLANK

ACKNOWLEDGMENTS

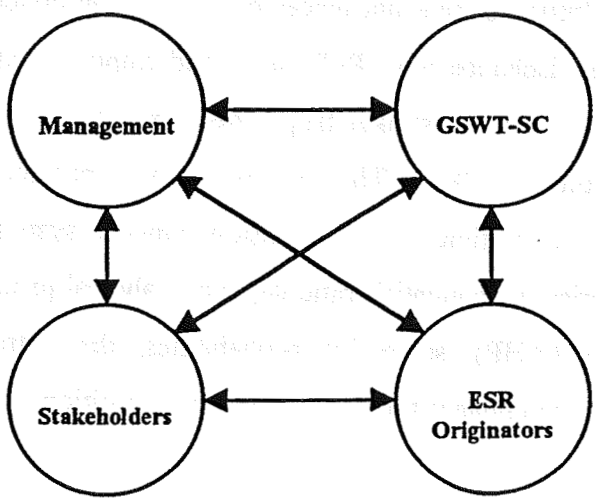
I would like to express my appreciation to NASA/ASEE for providing me with this wonderful research opportunity. I am greatly indebted to my NASA colleague, *Seunghee Lee* for her patience and expert guidance throughout this project. I m also grateful to *Jeff Wheeler* and *Bob Lang* for their support and encouragement. I would also like to thank the GSWT steering committee and all the members of stakeholder teams in Safety, Reliability, Systems Engineering, Operations, Programs Office, and PICB. And last but not least, I wish to express my appreciation to *Dr. E. Ramon Hosler*, Program Director, for his expert leadership and *Kari Stiles*, Administrative Assistant, for her professionalism and enthusiasm. They made participation in the program a pleasurable and rewarding experience.

ABSTRACT

The evaluation and prioritization of Engineering Support Requests (ESRs) is a particularly difficult task at the Kennedy Space Center (KSC) -- Shuttle Project Engineering Office. This difficulty is due to the complexities inherent in the evaluation process and the lack of structured information. The evaluation process must consider a multitude of relevant pieces of information concerning Safety, Supportability, O&M Cost Savings, Process Enhancement, Reliability, and Implementation. Various analytical and normative models developed over the past have helped decision makers at KSC utilize large volumes of information in the evaluation of ESRs. The purpose of this project is to build on the existing methodologies and develop a multiple criteria decision support system that captures the decision maker's beliefs through a series of sequential, rational, and analytical processes. The model utilizes the Analytic Hierarchy Process (AHP), subjective probabilities, the entropy concept, and Maximize Agreement Heuristic (MAH) to enhance the decision maker's intuition in evaluating a set of ESRs.

I. CONSENSUS RANKING APPROACH

Less available funding and the increasing number of ESRs has created more competition between stakeholders. There is a clear need to replace the current approach with a more comprehensive framework that promotes the participation and harmony among Management, GSWT Steering Committee, ESR Originators, and Stakeholders. [See Figure 1]



Organizational Benefits:

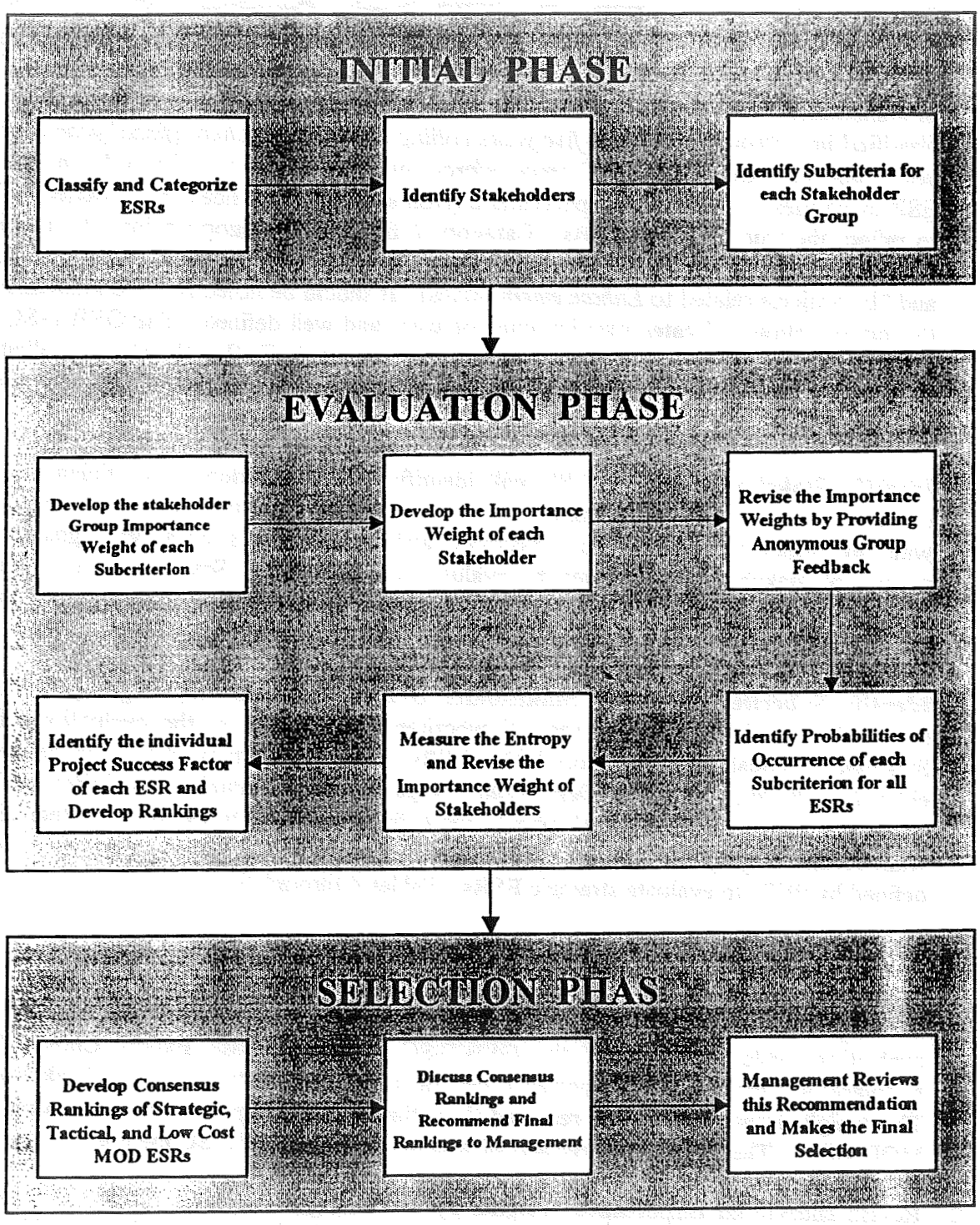
- Promotes Accountability
- Structured and Systematic
- Builds Consensus
- Enhances Participation and Satisfaction
- Mathematically Sound
- Easy to Implement
- Builds on the Current Framework [Tables 1 and 2]

Individual Threats:

The proposed framework requires a cultural change. Some GSWT Steering Committee members may perceive this framework threatening to their:

- Interests
- Authority
- Existence
- Unaccountability

Figure-1: Consensus Ranking Approach



INITIAL PHASE

1. **Classify and Categorize ESRs:** The GSWT Steering Committee (GSWT-SC) should identify different groups for classification and categorization of ESRs. Since ESRs come in different sizes (big budget items vs. medium or small) and scopes (long-term vs. short term or immediate), this grouping is needed for a meaningful comparison among them. ESRs are classified into *Strategic (three to five years rolling baseline)*, *Tactical (fiscal year to fiscal year)*, and *Low Cost MODS (Corrective/preventive maintenance)*. In addition, current ESR categories of 2S (Safety Impact) and 2 (Non-safety Impact) needs to be better defined to reflect the true nature of ESRs. Category 2 ESRs are categorized into 2O (projects related to *Obsolescence* issues), 2M (projects related to *High Maintenance/Design* issues), and 2E (projects related to *Enhancement* issues). It should be noted that the guidelines for this classification and categorization must be clear and well defined. The GSWT-SC will develop these guidelines and obtain management approval. GSWT-SC should solicit the ESR originator's opinion and justification concerning this classification and categorization. [Figures 2 and 3, Table 3]

2. **Identify Stakeholders:** GSWT-SC will identify the stakeholders to participate in the evaluation process and obtain management approval. This identification should be in line with the organizational mission, objectives, and management's fiscal year goals. Six groups of stakeholders identified to evaluate ESRs include: *Safety (Safety)*, *Systems Engineering (Supportability)*, *Program Office (O&M Cost Savings)*, *Operations (Process Enhancement)*, *Reliability (Reliability)*, and *PICB (Implementation)*. [Figure 4]

3. **Identify Subcriteria for each Stakeholder Group:** Each stakeholder group will meet separately and develops three sets of subcriteria to be used in the evaluation of the Strategic, Tactical, and Low Cost MOD ESRs. These subcriteria should be defined as either *Beneficial* (Positive) or *Detrimental* (Negative). For example, *Possibility of Death or Serious Injury* is a detrimental subcriterion defined by Safety to evaluate strategic ESRs while *Availability of Skilled Labor for Fixed-Priced Contract* is a beneficial subcriterion defined by PICB to evaluate strategic ESRs. [Tables 4 through 9]

EVALUATION PHASE

1. **Develop the Importance Weight of each Stakeholder:** At the beginning of each evaluation cycle, members of the GSWT-SC individually use Expert Choice (AHP software) and develop their importance weights of each evaluation criterion (stakeholder). This weight assessment will be repeated three times for Strategic, Tactical, and Low Cost MOD ESRs. The product is three sets of weights by each GSWT-SC member. [Table 10]

2. **Revise Individual Importance Weights by Providing Anonymous Group Feedback:** GSWT-SC meets and members review the anonymous feedback concerning individual and group weights. Members are encouraged to share their viewpoints and perceptions during this feedback session. At the end of the meeting, members can return and revise their weights, given their new insight and understanding from other individuals. The product is three sets of weights by each GSWT-SC member. [Table 11]

3. **Develop the Stakeholder Group Importance Weight of each Subcriterion:** Members of different stakeholder groups use Expert Choice in a brainstorming session and develop their group weight for each subcriterion (previously identified by the stakeholder groups). This step is repeated three times for Strategic, Tactical, and Low Cost MOD ESRs. The product is three sets of weights by each member of the stakeholder group. [Tables 4 through 9]

4. **Identify Probabilities of Occurrence of each Subcriterion for all ESRs:** Each stakeholder group receives a listing of all ESRs under consideration from the GSWT-SC. The stakeholder group will assign a probability to each subcriterion under each ESR. The assignment of probabilities could be done by the group in a brainstorming session or it could be done individually which can then be combined into a group judgment by calculating a Simple Mean. [Tables 12 through 15]

5. **Measure the Entropy and Revise the Importance Weight of Stakeholders:** Entropy concepts will be used to revise the initial weights of the subcriteria based on the information provided by the stakeholders concerning the probabilities. Given that each subcriterion is an information source, the more information is revealed by a subcriterion, the more relevant it is. This intrinsic information will be used in parallel with the stakeholder group weights. The probabilities of occurrence are used to measure this average intrinsic information. The more different the probabilities of a subcriteria are for a set of ESRs, the larger is the contrast intensity of the subcriterion and the greater is the amount of information transmitted by that subcriterion. [Tables 16 and 17]

6. **Identify the Individual Project Success Factor of each ESR and Develop Rankings:** The model will consolidate *importance weights of stakeholders* with the *weights for subcriteria* and the *probabilities of occurrence* to arrive at a set of Project Success Factors (PSF). The higher the PSF, the more desirable an ESR is. These calculations will be done separately for Strategic, Tactical, and Low Cost MOD ESRs. The product is three sets of individual rankings of the Strategic, Tactical, and Low Cost MOD ESRs. [Tables 18 through 21]

SELECTION PHASE

1. **Develop Consensus Rankings of Strategic, Tactical, and Low Cost MOD ESRs:** Maximize Agreement Heuristic (MAH) will be used to develop a Consensus Ranking of all Strategic, Tactical, and Low Cost MOD ESRs.

2. **Discuss Consensus Rankings and Recommend Final Rankings to Management:** The GSWT-SC meets and discusses the consensus rankings. A final recommendation that includes a ranking of all Strategic, Tactical, and Low Cost MOD ESRs will be forwarded to management for approval.

3. **Management Reviews this recommendation and makes the Final Selection:** Management reviews the recommendation of the GSWT-SC and makes the final Selection.

II. THE MATHEMATICAL MODEL

To formulate an algebraic model , let us assume:

V^m = Project Success factor of the m-th ESR; $(m = 1, 2, \ldots, q)$

W_i = The i-th Stakeholder Weight; $(i = 1, 2, \ldots, N_i)$

F_{ij} = The Overall Importance Weight for the j-th Subcriteria for the i-th Stakeholder; $(j = 1, 2, \ldots, N_j ; and\ i = 1, 2, \ldots, N_i)$

P_{ij}^m = The m-th Probability of Occurrence of the j-th Subcriteria for the i-th Stakeholder;$(m = 1, 2, \ldots, q; j = 1, 2, \ldots, N_j ; and\ i = 1, 2, \ldots, N_i)$

N_j = Number of Subcriteria for the i-th Stakeholder $(i = 1, 2, \ldots, N_i)$

N_i = Number of Stakeholders

Given that $i=1$ through N_i represents stakeholders, this study Has utilized 6 stakeholders including: Safety, Systems Engineering, Program office, Operations, Reliability, and PICB. The Project Success factor for the m-th ESR is:

$$V^m = \sum_{i=1}^{N_i} W_i \left(\sum_{j=1}^{Nj} F_{ij} \left(P_{ij}^m \right) \right)$$

Where:

$$\sum_{i=1}^{N_i} W_i = 1$$

$$\sum_{j=1}^{Nj} F_{ij} = 1$$

and

$$0 \le P_{ij}^m \le 1$$

III. THE CALCULATION OF ENTROPY

The model views decision making as an information processing task and a large amount of information about the strategic ESRs is processed through their subcriteria. Given the fact that subcriteria are information sources, the more information is revealed by the j-th subcriteria and the i-th stakeholder, the more relevant is the subcriteria in the decision analysis. Zeleny argues that this intrinsic information must be used in parallel with the initial weight assigned to various subcriteria by the DM. In other words, the overall importance weight of a subcriteria, F_{ij}, is directly related to the intrinsic weight, f_{ij}, reflecting average intrinsic information developed by a set of ESRs, and the subjective weight, W_i, reflecting the subjective assessment of its importance rendered by the DM. The probabilities of occurrence are used to measure this average intrinsic information. The more different the probabilities of a subcriteria are for a set of ESRs, the larger is the contrast intensity of the subcriteria, and the greater is the amount of information transmitted by that subcriteria. In this section, all formulas necessary for calculating the overall importance weight of opportunities are presented.

Assume a that vector $p_{ij} = (p_{ij}^1, \ldots, p_{ij}^q)$ characterizes the set P in terms of the j-th subcriteria in the i-th stakeholder and define:

$$P_{ij} = \sum_{m=1}^{q} p_{ij}^m \qquad \textit{(For } i = 1, 2, \ldots, N_i \textit{ and } j = 1, 2, \ldots, N_j\textit{)}$$

Then, the entropy measure of the j-th subcriteria in the i-th stakeholder is:

$$e(p_{ij}) = -K \sum_{m=1}^{q} \frac{p_{ij}^m}{P_{ij}} \ln \frac{p_{ij}^m}{P_{ij}}$$

Where $K > 0$, $\ln$ is the natural logarithm, $0 \le p_{ij}^m \le 1$, and $e(p_{ij}) \ge 0$. When all p_{ij}^m are equal for a given i and j, then $p_{ij}^m / P_{ij} = 1/q$, and $e(p_{ij})$ assumes its maximum value, which is $e_{\max} = \ln q$. By setting $K = 1/e_{\max}$, we achieve $0 \le e(p_{ij}) \le 1$. This normalization is necessary for meaningful comparisons. In addition, the total entropy is defined as:

$$E = \sum_{j=1}^{N_j} e(p_{ij})$$

The smaller $e(p_{ij})$ is, the more information is transmitted by the j-th subcriteria in the i-th stakeholder and the larger $e(p_{ij})$, the less information is transmitted. When $e(p_{ij}) = e_{max} = \ln q$, the j-th subcriteria in the i-th stakeholder is not transmitting any useful information. Next, the intrinsic weight is calculated as:

$$f_{ij} = \frac{1}{N_i - E}\left[1 - e(p_{ij})\right]$$

Since f_{ij} is inversely related to $e(p_{ij})$, $1 - e(p_{ij})$ is used instead of $e(p_{ij})$ and normalized to make sure $0 \le f_{ij} \le 1$ and

$$\sum_{j=1}^{N_J} f_{ij} = 1$$

The more different the subjective probabilities, p_{ij}^m, are, the larger f_{ij}, and the more important the j-th subcriteria in the i-th stakeholder is. When all the subjective probabilities, p_{ij}^m, are equal, then $f_{ij} = 0$. In order to calculate the overall importance weight of the j-th subcriteria in the i-th stakeholder, F_{ij}, the intrinsic weight, f_{ij}, is multiplied by the subjective weight, w_{ij}, and then the product is normalized:

$$F_{ij} = \frac{f_{ij} \cdot w_{ij}}{\sum_{j=1}^{N_J} f_{ij} \cdot w_{ij}}$$

IV. THE ANALYTIC HIERARCHY PROCESS

Measuring the Relative Importance of *n* Factors:

Assume that in a decision maker's (DM's) mind, c_1, c_2,...,c_n are the n (Level 1) factors that contribute to an ESR success (Level 0). The DM's goal is to assess the relative importance of these factors. Saaty's Analytic Hierarchy Process (AHP) is a method of deriving a set of weights to be associated with each of the n factors, and it works as below:

The DM is asked to compare each possible pair c_i, c_j of factors at a given level, and provide quantified judgments on which one of the factors is more important and by how much. These judgments are represented by an n x n matrix:

$$A = (a_{ij}) \qquad (i, j = 1, 2, 3....n)$$

If c_i is judged to be of equal importance as c_j, then $a_{ij}=1$

If c_i is judged to be more important than c_j, then $a_{ij}>1$

If c_i is judged to be less important than c_j, then $a_{ij}<1$

$$a_{ij} = 1/a_{ji} \qquad a_{ij} \neq 0$$

Thus, the matrix A is a reciprocal matrix (i.e., the entry a_{ij} is the inverse of the entry a_{ji}). a_{ij} reflects the relative importance of c_i compared with factor c_j. For example, $a_{12}=1.25$ indicates that c_1 is 1.25 times as important as c_2.

Then, the vector w representing the relative weights of each of the n factors can be found by computing the normalized eigenvector corresponding to the maximum eigenvalue of matrix A. An eigenvalue of A is defined as λ which satisfies the following matrix equation:

$$A\,w = \lambda\,w$$

where λ is a constant, called the eigenvalue, associated with the given eigenvector w. Saaty has shown that the best estimate of w is the one associated with the maximum eigenvalue (λ_{max}) of the matrix A. Since the sum of the weights should be equal to 1.00, the normalized eigenvector is used. Saaty's algorithm for obtaining this w is incorporated in the software Expert Choice.

Measuring Consistency:

One of the advantages of AHP is that it ensures that DMs are consistent in their pairwise comparisons. Saaty suggests a measure of consistency for the pairwise comparisons. When the judgments are perfectly consistent, the maximum eigenvalue, λ_{max}, should equal n, the number of factors that are compared. In general, the responses are not perfectly consistent, and λ_{max} is greater than n. The larger the λ_{max}, the greater is the degree of inconsistency. Saaty defines the consistency index (*CI*) as $(\lambda_{max} - n) / (n - 1)$, and provides the following random index (*RI*) table for matrices of order *3* to *10*. This *RI* is based on a simulation of a large number of randomly generated weights. Saaty recommends the calculation of a consistency ratio (*CR*), which is the ratio of *CI* to the *RI* for the same order matrix. A *CR* of *0.10* or less is considered acceptable. When the *CR* is unacceptable, the DM is made aware that his or her pairwise comparisons are logically inconsistent, and he or she is encouraged to revise the same.

n	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
RI	*0.58*	*0.90*	*1.12*	*1.32*	*1.41*	*1.45*	*1.49*	*1.51*

Computation of the Preference Rating for Each Criteria:

In the simplest application of AHP, the problem is decomposed into *3* levels. The top level (Level 0) consists of one factor which is the overall goal. The second level consists of n factors or attributes that define the overall goal. The third level consists of the m choices that are to be evaluated.

Let v_i represent the normalized weight given to the i-th factor relative to the overall goal. Let w_{ij} represent the normalized preference given to the i-th criteria with respect to the j-th factor. Then the preference rating (p_i) of the i-th choice with respect to the overall goal is calculated as:

$$P_i = \sum_{j=1}^{n} v_j w_{ij}$$

In the standard practice of AHP, when more than one DM is involved, the linear composite concept is extended across various DMs. In this project the Maximize Agreement Heuristic is used to aggregate individual GSWT committee member preferences.

V. MAXIMIZE AGREEMENT HEURISTIC (MAH)

Assume that each one of a group of k decision makers (DM) has ranked n ESRs. Assuming further that the opinions of the k DMs are to be valued equally, the Maximize Agreement Heuristic (MAH) seeks to arrive at the consensus ranking of the ESRs for the group as a whole. According to Beck and Lin, MAH defines an agreement matrix, A, where each element a_{ij} represents the number of DMs who have preferred ESR i to ESR j. Strict preference is important. If a DM is indifferent between i and j, he or she is not counted in a_{ij}. The sum of a_{ij} for each ESR i across all columns represents the positive preference vector, P, where

$$P_i = \sum_{j=1}^{n} a_{ij}, \qquad i=1,2,3,\ldots,n$$

Similarly, the sum of a_{ij} for each ESR across all rows represents the negative preference vector, N, where

$$N_i = \sum_{j=1}^{n} a_{ji}, \qquad i=1,2,3,\ldots,n$$

If for ESR i, $P_i = 0$, implying that no DM prefers ESR i to any other ESR, ESR i is placed at the bottom [in subsequent iterations, at the next available position at the bottom] of the final consensus ranking. However, if for ESR i, $N_i = 0$, implying that no DM prefers any other ESR over ESR i, ESR i is placed at the top [in subsequent iterations, at the next available position at the top] of the ranking.

When there are no zero values in either P or N, the difference in total decision maker agreement and disagreement ($P_i - N_i$) is calculated for each ESR, and ESR i with the largest absolute difference $|P_i - N_i|$ is considered. If ($P_i - N_i$) is positive, ESR i is placed in the next available position at top of the final consensus ranking, and if the difference is negative, ESR i is placed in the next available position at the bottom of the consensus ranking. Any ties are broken arbitrarily. Once a ESR is assigned a position in the final consensus ranking, that ESR is eliminated from further consideration. The remaining ESRs form a new matrix and the process is repeated until all ESRs are ranked.

Table -1: Current Approach

Stakeholders	Weight	TE K15694	TE K16038	TE K1573657	TE K15470	TE K15474	TE OB32-56	TE K16055	TE K15998	TE K16052	TE OB35-100	TE K15660	TE K16053	TE FP017-2A	TE OB35-06	TE K16056	TE K15816	TE K16045	TE K16012	TE K15787	TE K15659	TE K16054	TE K16006
Safety	0.320	4	6	9	10	10	1	3	9	5	5	6	5	4	5	3	2	10	1	2	8	1	1
Systems Engineering	0.200	2	9	9	0	0	8	1	3	0	2	3	0	5	1	2	5	2	2	8	2	5	8
Program Office	0.170	10	7	8	9	4	6	5	5	8	5	9	8	4	9	5	8	6	9	7	7	8	9
Operations	0.160	8	5	6	4	6	4	6	4	9	5	8	9	7	5	10	5	5	4	9	9	7	3
Reliability	0.100	3	4	8	1	4	5	1	1	1	6	1	1	7	1	1	10	4	1	4	1	3	3
PICB	0.050	3	4	5	2	0	2	2	6	8	4	6	9	8	5	5	8	6	7	7	5	8	3
OVERALL SCORE		5.11	6.31	8.05	5.57	5.24	4.18	3.17	5.37	4.90	4.45	5.73	4.95	5.18	4.48	4.16	5.20	6.12	3.34	5.62	5.94	4.50	4.38

Stakeholders	Weight	TV K13747	TV K16026	TV K13858	TV K14887	TV K13643	TV K13983	TV K14513	TV K13512	TV K15697	TV K13936	TV K15684	TV K15842	TV K15504	TV K15533	TV K15770	TV K15903	TV K14920	TV K13562	TV K15569
Safety	0.320	7	1	1	1	1	4	5	1	8	9	7	10	9	1	2	1	1	5	1
Systems Engineering	0.200	0	0	5	7	8	8	5	0	0	2	10	9	1	3	3	4	4	2	5
Program Office	0.170	7	9	9	8	8	6	5	7	8	5	5	8	6	6	4	6	5	8	7
Operations	0.160	5	0	9	8	8	5	0	9	0	4	6	7	7	9	1	2	3	2	0
Reliability	0.100	1	1	6	8	5	3	1	1	1	1	1	10	4	3	1	4	1	1	2
PICB	0.050	2	7	10	7	5	4	6	6	8	5	9	8	6	7	6	7	8	6	4
OVERALL SCORE		4.43	2.30	5.39	5.51	5.31	5.20	3.85	3.35	4.42	5.12	6.60	8.88	5.92	4.03	2.48	3.21	2.95	4.08	2.91

Note: The Operations scores of the TV projects are tentative and subject to the approval of the directors.

GSWT Steering Committee Identifies the Stakeholders and Provides their Importance Weights.
Stakeholders Provide Subjective Scores between 1 through 10 for each ESR.

Table-2: ESR Rankings Using the Current Approach

RANK	ESR Number	Organization	Category	ROM	Overall Score
1	K15642	TV	2S	17,077	8.88
2	K15736/37	TE	2S	175,060	8.05
3	K15684	TV	2S	13,785	6.60
4	K16038	TE	2	56,600	6.31
5	K16045	TE	2S	20,200	6.12
6	K15659	TE	2S	30,000	5.94
7	K15504	TV	2S	64,383	5.92
8	K15660	TE	2	15,000	5.73
9	K15787	TE	2	46,000	5.62
10	K15470	TE	2S	82,263	5.57
11	K14887	TV	2	20,680	5.51
12	K15858	TV	2	55,000	5.39
13	K15998	TE	2S	112,000	5.37
14	K15645	TV	2	26,942	5.31
15	K15674	TE	2S	18,000	5.24
16	K15616	TE	2	46,600	5.20
17	K15985	TV	2	49,200	5.20
18	FF017-2A	TE	2	50,000	5.18
19	K15936	TV	2S	29,000	5.12
20	K15694	TE	2	5,700	5.11
21	K16033	TE	2S	22,000	4.95
22	K16032	TE	2S	16,000	4.90
23	K16036	TE	2	14,200	4.50
24	GS35-86	TE	2	7,200	4.48
25	GS35-100	TE	2	70,000	4.45
26	K15747	TV	2S	240,029	4.43
27	K15697	TV	2S	20,196	4.42
28	K16000	TE	2	35,000	4.38
29	GS33-56	TE	2	50,000	4.18
30	K16006	TE	2	14,400	4.16
31	K15562	TV	2	34,830	4.08
32	K15535	TV	2	110,662	4.03
33	K14515	TV	2	117,647	3.85
34	K15512	TV	2	19,316	3.35
35	K16012	TE	2S	40,000	3.34
36	K15563	TV	2	1,850	3.21
37	K16055	TE	2	29,000	3.17
38	K14920	TV	2	111,500	2.95
39	K15569	TV	2	61,570	2.91
40	K15770	TV	2	2,000	2.48
41	K16026	TV	2	8,000	2.30

Figure-2: Classify and Categorize ESRs (Initial Phase)

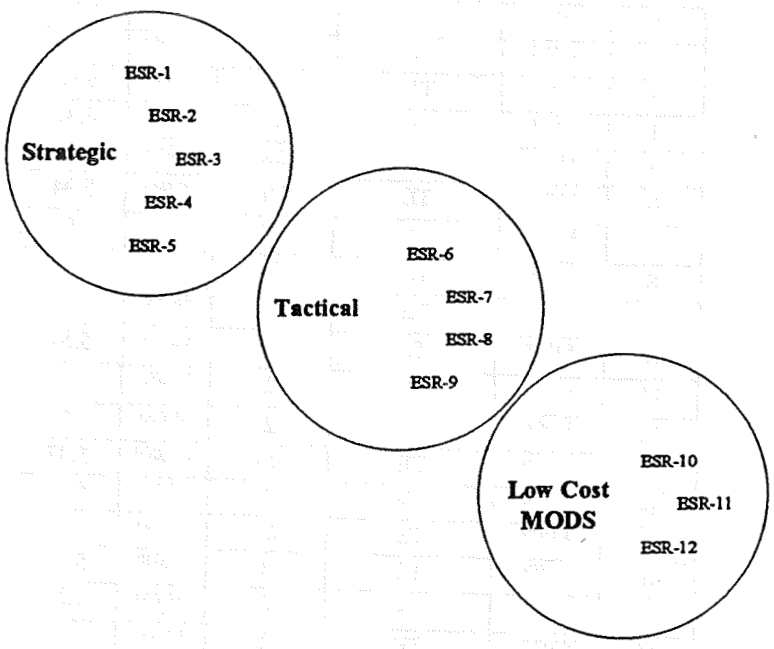


Table-3: ESR Classifications and Their Distinctive Characteristics

Order of Importance	Distinctive Characteristics	Strategic	Tactical	Low Cost MODS
1	Benefits	Program	Mission	Operational
2	Urgency	Future Start	Start This Year	Immediate Start
3	Requested Budget ($) (Material+Labor)	$ > 250,000	LCM ≤ $ ≤ 250,000	$ < 5,000 100 Man Hours Design 100 Man Hours Labor
4	Technology	Developmental	Adaptable	Off-the-shelf
5	Metrics	Long-Term Fix	Short-Term Fix	Immediate Fix
6	Implementation (Time)	Multi Year	Single Year	Under One Year

Figure-3: Categorize Non-Safety Impact ESRs into 2O, 2M, and 2E

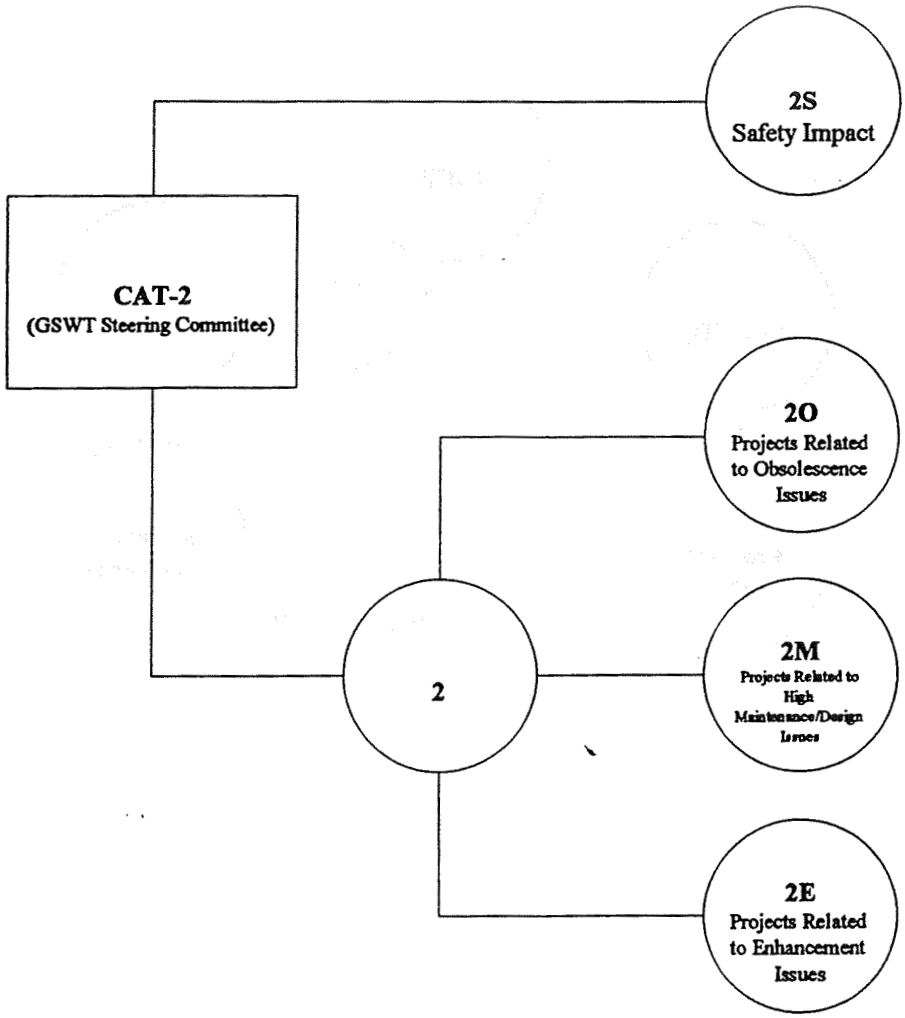


Figure 4: Identify Stakeholders (Initial Phase)

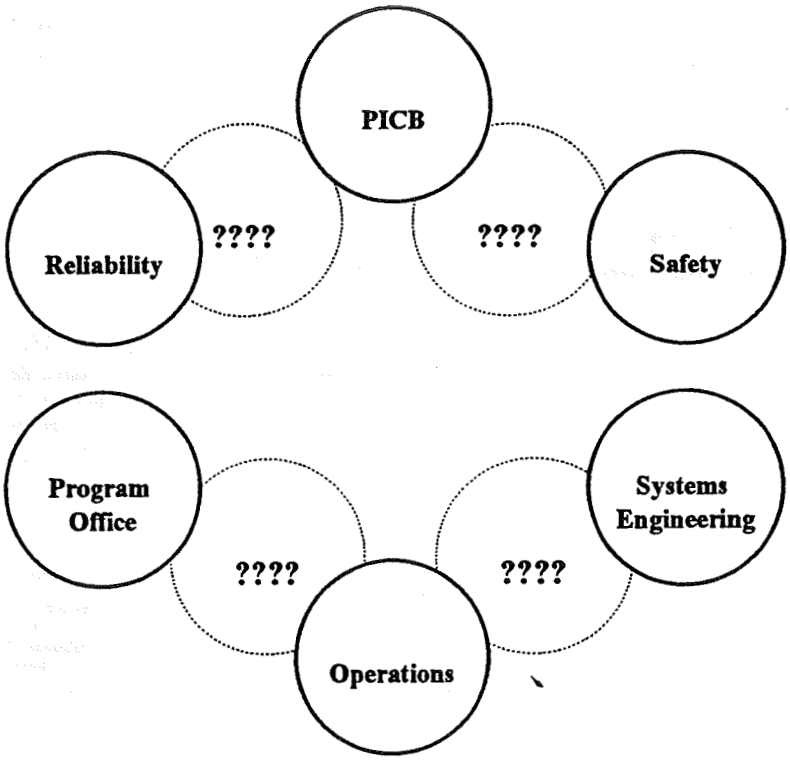


Table-4: Identify Subcriteria and Their Importance Weights for each Stakeholder Group
STAKEHOLDER: Safety -- CRITERIA: Safety

Impact	Sub-code	Strategic Subcriteria	Weight
-1	SS-DSI	Possibility of Death or Serious Injury	0.564
-1	SS-LOF	Possibility of Loss of Flight Hardware, Facility, or GSE	0.239
-1	SS-PID	Possibility of Personal Injury and/or Flight Hardware, Facility, or GSE Damage	0.118
-1	SS-SVS	Possibility of a Serious Violation of Safety, Health, or Environmental Federal/State Regulation	0.047
-1	SS-DVS	Possibility of a Deminius Violation of Safety, Health, or Environmental Federal/State Regulation	0.032

Impact	Sub-code	Tactical Subcriteria	Weight
-1	ST-DSI	Possibility of Death or Serious Injury	0.564
-1	ST-LOF	Possibility of Loss of Flight Hardware, Facility, or GSE	0.239
-1	ST-PID	Possibility of Personal Injury and/or Flight Hardware, Facility, or GSE Damage	0.118
-1	ST-SVS	Possibility of a Serious Violation of Safety, Health, or Environmental Federal/State Regulation	0.047
-1	ST-DVS	Possibility of a Deminius Violation of Safety, Health, or Environmental Federal/State Regulation	0.032

Impact	Sub-code	Low Cost MODS Subcriteria	Weight
-1	SL-DSI	Possibility of Death or Serious Injury	0.564
-1	SL-LOF	Possibility of Loss of Flight Hardware, Facility, or GSE	0.239
-1	SL-PID	Possibility of Personal Injury and/or Flight Hardware, Facility, or GSE Damage	0.118
-1	SL-SVS	Possibility of a Serious Violation of Safety, Health, or Environmental Federal/State Regulation	0.047
-1	SL-DVS	Possibility of a Deminius Violation of Safety, Health, or Environmental Federal/State Regulation	0.032

Table-5: Identify Subcriteria and Their Importance Weights for each Stakeholder Group

STAKEHOLDER: Systems Engineering -- CRITERIA: Supportability

Impact	Sub-code	Strategic Subcriteria	Weight
1	ES-REQ	Ability to Support Program Requirements for the Balance of the Program's Life	0.444
-1	ES-NON	Occurrence of Non-Support Activities	0.237
1	ES-TCH	Ability to Eliminate Reliance on Expected Obsolete Technology	0.174
1	ES-MSP	Ability to Reduce the Probability of Milestone Slippage	0.083
1	ES-RFR	Ability to Reduce Failure Rate	0.062

Impact	Sub-code	Tactical Subcriteria	Weight
1	ET-LSP	Ability to Reduce the Probability of Launch Slippage	0.553
1	ET-NTR	Ability to Support Program for Near-Term Requirements	0.171
-1	ET-NON	Occurrence of Non-Support Activities	0.132
1	ET-FIX	Ability to Fix a Failure	0.107
1	ET-TCH	Ability to Eliminate Reliance on Identified Obsolete Technology	0.037

Impact	Sub-code	Low Cost MOD Subcriteria	Weight
-1	EL-NON	Occurrence of Non-Support Activities	0.285
1	EL-ITR	Ability to Support Program for Immediate Requirements	0.282
1	EL-FIX	Ability to Fix a Failure	0.227
1	EL-PSS	Ability to Reduce the Probability of Process and Schedule Slippage	0.165
1	EL-TCH	Ability to Eliminate Reliance on Identified Obsolete Technology	0.041

Table-6: Identify Subcriteria and Their Importance Weights for each Stakeholder Group
STAKEHOLDER: Program Office -- CRITERIA: O&M Cost Savings

Impact	Sub-code	Strategic Subcriteria	Weight
1	PS-CON	Ability to Meet Contractual Obligations	0.370
1	PS-CST	Ability to Meet the Proposed Cost	0.280
1	PS-OMC	Reduction of O&M Costs	0.179
1	PS-PAY	Payback Period with Program Duration	0.087
1	PS-IMP	Ability to Utilize Existing Workforce During Implementation	0.044
1	PS-SCH	Ability to Meet the Proposed Schedule	0.040

Impact	Sub-code	Tactical Subcriteria	Weight
1	PT-PRI	Ability to Meet Safety/Launch & Landing Criteria	0.391
1	PT-FUN	Availability of Funds	0.197
1	PT-IMP	Ability to Utilize Time-Sensitive Implementation Methodology	0.147
1	PT-CST	Ability to Meet the Proposed Cost	0.105
1	PT-SCH	Ability to Meet the Proposed Schedule	0.086
1	PT-OMC	Reduction of O&M Costs	0.045
1	PT-CON	Ability to Meet Contractual Obligations	0.029

Impact	Sub-code	Low Cost MODS Subcriteria	Weight
-1	PL-EFF	Potential "Domino" Effect on Systems or Facilities	0.714
-1	PL-DES	Possibility of Design Change Requirements	0.143
1	PL-QCK	Ability to Turn-around the Project Quick	0.143

Table-7: Identify Subcriteria and Their Importance Weights for each Stakeholder Group
STAKEHOLDER: Operations -- CRITERIA: Process Enhancement
Ground Engineering (TE)

Impact	Sub-code	Strategic Subcriteria	Weight
1	OS-PEO	TE: Ability to Use Less People	0.661
1	OS-TSK	TE: Ability to Simplify/Eliminate Tasks	0.208
1	OS-TIM	TE: Ability to Reduce Time	0.131

Impact	Sub-code	Tactical Subcriteria	Weight
1	OT-PEO	TE: Ability to Use Less People	0.661
1	OT-TSK	TE: Ability to Simplify/Eliminate Tasks	0.208
1	OT-TIM	TE: Ability to Reduce Time	0.131

Impact	Sub-code	Low Cost MODS Subcriteria	Weight
1	OL-PEO	TE: Ability to Use Less People	0.661
1	OL-TSK	TE: Ability to Simplify/Eliminate Tasks	0.208
1	OL-TIM	TE: Ability to Reduce Time	0.131

Vehicle Engineering (TV)

Impact	Sub-code	Strategic Subcriteria	Weight
1	OS-PEP	TV: Ability to Use Less People	0.577
1	OS-TIM	TV: Ability to Reduce Time	0.342
1	OS-LOC	TV: Accessibility of the Work Location	0.081

Impact	Sub-code	Tactical Subcriteria	Weight
1	OT-PEP	TV: Ability to Use Less People	0.563
1	OT-TIM	TV: Ability to Reduce Time	0.246
1	OT-LOC	TV: Ability to Access the Work Location	0.124
1	OT-HNM	TV: Ability to Reduce/Eliminate Hardware/Materials Expended During Processing	0.067

Impact	Sub-code	Low Cost MODS Subcriteria	Weight
1	OL-PEP	TV: Ability to Use Less People	0.563
1	OL-TIM	TV: Ability to Reduce Time	0.246
1	OL-LOC	TV: Ability to Access the Work Location	0.124
1	OL-HNM	TV: Ability to Reduce/Eliminate Hardware/Materials Expended During Processing	0.067

Table-8: Identify Subcriteria and Their Importance Weights for each Stakeholder Group

STAKEHOLDER: Reliability -- CRITERIA: Reliability

Impact	Sub-code	Strategic Subcriteria	Weight
1	RS-SFP	Eliminate Critical Single Failure Points (CSFPs)	0.412
1	RS-PFP	Reduce the Possibility of Failure Propagation to other Components or Systems	0.194
1	RS-MTR	Improve Mean Time to Repair (MTTR)	0.110
1	RS-FII	Improve Fault Identification/Fault Isolation (FI/FI)	0.092
1	RS-SIM	Provide for a Simpler System	0.053
1	RS-AMT	Improve Access for Maintenance Tasks	0.049
1	RS-TBF	Increase Mean Time Between Failures (MTBFs)	0.040
1	RS-ETT	Reduce Support Equipment, Special Tools, and Special Training Requirements	0.030
1	RS-COT	Provide for the use of Standard Commercial of-the-shelf (COTS) Parts	0.010
1	RS-EQI	Provide for Equipment Interchangability	0.010

Impact	Sub-code	Tactical Subcriteria	Weight
1	RT-SFP	Eliminate Critical Single Failure Points (CSFPs)	0.412
1	RT-PFP	Reduce the Possibility of Failure Propagation to other Components or Systems	0.194
1	RT-MTR	Improve Mean Time to Repair (MTTR)	0.110
1	RT-FII	Improve Fault Identification/Fault Isolation (FI/FI)	0.092
1	RT-SIM	Provide for a Simpler System	0.053
1	RT-AMT	Improve Access for Maintenance Tasks	0.049
1	RT-TBF	Increase Mean Time Between Failures (MTBFs)	0.040
1	RT-ETT	Reduce Support Equipment, Special Tools, and Special Training Requirements	0.030
1	RT-COT	Provide for the use of Standard Commercial of-the-shelf (COTS) Parts	0.010
1	RT-EQI	Provide for Equipment Interchangability	0.010

Impact	Sub-code	Low Cost MODS Subcriteria	Weight
1	RL-SFP	Eliminate Critical Single Failure Points (CSFPs)	0.412
1	RL-PFP	Reduce the Possibility of Failure Propagation to other Components or Systems	0.194
1	RL-MTR	Improve Mean Time to Repair (MTTR)	0.110
1	RL-FII	Improve Fault Identification/Fault Isolation (FI/FI)	0.092
1	RL-SIM	Provide for a Simpler System	0.053
1	RL-AMT	Improve Access for Maintenance Tasks	0.049
1	RL-TBF	Increase Mean Time Between Failures (MTBFs)	0.040
1	RL-ETT	Reduce Support Equipment, Special Tools, and Special Training Requirements	0.030
1	RL-COT	Provide for the use of Standard Commercial of-the-shelf (COTS) Parts	0.010
1	RL-EQI	Provide for Equipment Interchangability	0.010

Table-9: Identify Subcriteria and Their Importance Weights for each Stakeholder Group
STAKEHOLDER: PICB -- CRITERIA: Implementation

Impact	Sub-code	Strategic Subcriteria	Weight
1	IS-FPC	Fixed-Priced Contract Implementation Versus In-house Labor	0.446
-1	IS-FMC	Possibility of Flight Manifest Changes	0.193
1	IS-PRC	Ability to Meet Program-Related Changes	0.122
1	IS-TCH	Ability to Meet New Technology Considerations	0.108
1	IS-COF	Sufficiency of Funds to Push ESR to Construction of Facility (C of F) Funding	0.047
1	IS-EPA	Ability to Meet all Environmental Considerations and Regulations (EPA)	0.045
-1	IS-WOO	Possibility of Interference in Implementation (Window of Opportunity)	0.039

Impact	Sub-code	Tactical Subcriteria	Weight
-1	IT-MSA	Multi-Site Applicability	0.423
-1	IT-WOO	Possibility of Interference in Implementation (Window of Opportunity)	0.195
-1	IT-FMC	Possibility of Flight Manifest Changes	0.137
-1	IT-MSC	Effects on Multi-System Configuration Systems	0.116
-1	IT-EOH	Possibility of Equipment and Occupational Hazards	0.065
-1	IT-SSR	Site Specific Restrictions	0.033
1	IT-TCH	Ability to Meet New Technology Considerations	0.031

Impact	Sub-code	Low Cost MODS Subcriteria	Weight
1	IL-OBJ	ESR Alignment with Organizational Goals and objectives	0.436
-1	IL-EOH	Possibility of Equipment and Occupational Hazards	0.318
-1	IL-FMC	Possibility of Flight Manifest Changes	0.097
1	IL-TCH	Ability to Meet New Technology Considerations	0.076
1	IL-ALT	Availability of Alternative Solutions/Feasible Work Around	0.073

Table-10: Develop the Importance Weight of each Stakeholder (Round 1)

GSWT Committee Member	Safety	Supportability	O&M Cost Savings	Process Enhancement	Reliability	Implementation	IR
A11	0.450	0.256	0.168	0.042	0.055	0.029	0.070
B12	0.426	0.234	0.045	0.038	0.217	0.040	0.080
C13	0.492	0.051	0.144	0.040	0.202	0.071	0.090
D14	0.526	0.140	0.135	0.136	0.034	0.029	0.060
E15	0.534	0.187	0.085	0.042	0.114	0.038	0.070
F16	0.340	0.075	0.129	0.028	0.226	0.202	0.080
G17	0.583	0.104	0.093	0.050	0.140	0.030	0.070
H18	0.478	0.073	0.100	0.113	0.205	0.031	0.060
I19	0.480	0.243	0.135	0.077	0.038	0.027	0.070
J20	0.434	0.116	0.065	0.043	0.308	0.034	0.090
K21	0.436	0.125	0.090	0.065	0.241	0.043	0.070
L22	0.382	0.165	0.036	0.113	0.265	0.039	0.070
M23	0.506	0.149	0.039	0.039	0.203	0.064	0.090
N24	0.392	0.268	0.071	0.074	0.162	0.033	0.080
O25	0.460	0.031	0.162	0.140	0.170	0.037	0.080
P26	0.187	0.243	0.052	0.287	0.142	0.089	0.070
Geometric Mean	0.432	0.131	0.087	0.067	0.144	0.044	
Normalized Mean	**0.477**	**0.145**	**0.096**	**0.074**	**0.160**	**0.048**	
Current Weights	**0.320**	**0.200**	**0.170**	**0.160**	**0.100**	**0.050**	
Absolute Deviation	**0.157**	**0.055**	**0.074**	**0.086**	**0.060**	**0.002**	

Table-11: Develop the Importance Weight of each Stakeholder (Round 2)

GSWT Committee Member	Safety	Supportability	O&M Cost Savings	Process Enhancement	Reliability	Implementation	IR
A11	0.119	0.262	0.242	0.139	0.194	0.044	0.080
B12	0.417	0.274	0.043	0.086	0.151	0.029	0.080
C13	0.439	0.052	0.165	0.049	0.218	0.077	0.080
D14	0.515	0.146	0.134	0.134	0.045	0.026	0.060
E15	0.534	0.187	0.085	0.042	0.114	0.038	0.070
F16	0.374	0.234	0.141	0.060	0.111	0.080	0.050
G17	0.383	0.151	0.166	0.048	0.217	0.035	0.050
H18	0.478	0.073	0.100	0.113	0.205	0.031	0.060
I19	0.437	0.220	0.034	0.056	0.150	0.103	0.070
J20	0.440	0.114	0.072	0.035	0.309	0.030	0.060
L22	0.498	0.147	0.061	0.063	0.201	0.030	0.080
M23	0.431	0.148	0.052	0.035	0.294	0.040	0.060
N24	0.392	0.268	0.071	0.074	0.162	0.033	0.080
O25	0.480	0.056	0.241	0.094	0.100	0.029	0.070
P26	0.187	0.243	0.052	0.287	0.142	0.089	0.070
Geometric Mean	0.385	0.152	0.093	0.073	0.159	0.042	
Normalized Mean	**0.426**	**0.168**	**0.103**	**0.081**	**0.176**	**0.047**	
Current Weights	**0.320**	**0.200**	**0.170**	**0.160**	**0.100**	**0.050**	
Absolute Deviation	**0.106**	**0.032**	**0.067**	**0.079**	**0.076**	**0.003**	
First Round Weights	**0.477**	**0.145**	**0.096**	**0.074**	**0.160**	**0.048**	
Second Round Weights	**0.426**	**0.168**	**0.103**	**0.081**	**0.176**	**0.047**	
Absolute Deviation	**0.052**	**0.023**	**0.007**	**0.007**	**0.016**	**0.001**	

Table-12: Identify Probabilities of Occurrence of Each Subcriterion for all ESRs and Measure the Entropy and Revise the Importance Weight of Stakeholders

Stakeholders	Criteria	GSWT Steering Committee															Mean
		J. Alexander	M. Allison	G. Fischer	J. Borchell	D. Kelley	C. Ost	S. Lee	K. Mellott	D. Haisen	S. Hunter	M. Steele	R. Tetrall	P. Serpkins	J. White	K. Kelsey	J. Wemmit	
Safety	Safety	0.119	0.417	0.439	0.515	0.534	0.374	0.383	0.478	0.437	0.440	0.436	0.498	0.431	0.392	0.480	0.187	0.426
Systems Engineering	Supportability	0.262	0.274	0.052	0.146	0.187	0.234	0.151	0.073	0.220	0.114	0.125	0.147	0.148	0.268	0.056	0.243	0.168
Program Office	O&M Cost Savings	0.242	0.043	0.165	0.134	0.085	0.141	0.166	0.100	0.034	0.072	0.090	0.061	0.052	0.071	0.241	0.052	0.103
Operations	Process Enhancement	0.139	0.086	0.049	0.134	0.042	0.060	0.048	0.113	0.056	0.035	0.065	0.063	0.035	0.074	0.094	0.287	0.081
Reliability	Reliability	0.194	0.151	0.218	0.045	0.114	0.111	0.217	0.205	0.150	0.309	0.241	0.201	0.294	0.162	0.100	0.142	0.176
PICB	Implementation	0.044	0.029	0.077	0.026	0.038	0.080	0.035	0.031	0.103	0.030	0.043	0.030	0.040	0.033	0.029	0.089	0.046

Table-13: Identify Probabilities of Occurrence of Each Subcriterion for all ESRs and Measure the Entropy and Revise the Importance Weight of Stakeholders

Stakeholders	Impact	Sub-code	Subcriteria (Tactical ESRs)	Imp. Weight	Intrinsic Weight	Overall Weight
Safety	-1	ST-DSI	Possibility of Death or Serious Injury	0.564	0.451	0.762
	-1	ST-LOF	Possibility of Loss of Flight Hardware, Facility, or GSE	0.239	0.204	0.146
	-1	ST-PID	Possibility of Personal Injury and/or Flight Hardware, Facility, or GSE Damage	0.118	0.222	0.079
	-1	ST-SVS	Possibility of a Serious Violation of Safety, Health, or Environmental Federal/State Regulation	0.047	0.040	0.006
	-1	ST-DVS	Possibility of a Deminius Violation of Safety, Health, or Environmental Federal/State Regulation	0.032	0.083	0.008
Systems Engineering	+1	ET-LSP	Ability to Reduce the Probability of Launch Slippage	0.553	0.285	0.736
	+1	ET-NTR	Ability to Support Program for Near-Term Requirements	0.171	0.121	0.097
	-1	ET-NON	Occurrence of Non-Support Activities	0.132	0.042	0.026
	+1	ET-FIX	Ability to Fix a Failure	0.107	0.140	0.070
	+1	ET-TCH	Ability to Eliminate Reliance on Identified Obsolete Technology	0.037	0.411	0.071
Program Office	+1	PT-PRI	Ability to Meet Safety/Launch & Landing Criteria	0.391	0.351	0.743
	+1	PT-FUN	Availability of Funds	0.197	0.058	0.062
	+1	PT-IMP	Ability to Utilize Time-Sensitive Implementation Methodology	0.147	0.035	0.028
	+1	PT-CST	Ability to Meet the Proposed Cost	0.105	0.038	0.021
	+1	PT-SCH	Ability to Meet the Proposed Schedule	0.086	0.086	0.040
	+1	PT-OMC	Reduction of O&M Costs	0.045	0.427	0.104
	+1	PT-CON	Ability to Meet Contractual Obligations	0.029	0.004	0.001
Operations	+1	OT-PEO	TE: Ability to Use Less People	0.661	0.572	0.832
	+1	OT-TSK	TE: Ability to Simplify/Eliminate Tasks	0.208	0.264	0.121
	+1	OT-TIM	TE: Ability to Reduce Time	0.131	0.163	0.047
	+1	OT-PEP	TV: Ability to Use Less People	0.563	0.227	0.578
	+1	OT-TIM	TV: Ability to Reduce Time	0.246	0.129	0.144
	+1	OT-LOC	TV: Ability to Access the Work Location	0.124	0.321	0.180
	+1	OT-HNM	TV: Ability to Reduce/Eliminate Hardware/Materials Expended During Processing	0.067	0.323	0.098
Reliability	+1	RT-SFP	Eliminate Critical Single Failure Points (CSFPs)	0.412	0.213	0.634
	+1	RT-PFP	Reduce the Possibility of Failure Propagation to other Components or Systems	0.194	0.097	0.136
	+1	RT-MTR	Improve Mean Time to Repair (MTTR)	0.110	0.067	0.054
	+1	RT-FII	Improve Fault Identification/Fault Isolation (FI/FI)	0.092	0.072	0.048
	+1	RT-SIM	Provide for a Simpler System	0.053	0.109	0.042
	+1	RT-AMT	Improve Access for Maintenance Tasks	0.049	0.094	0.033
	+1	RT-TBF	Increase Mean Time Between Failures (MTBFs)	0.040	0.076	0.022
	+1	RT-ETT	Reduce Support Equipment, Special Tools, and Special Training Requirements	0.030	0.081	0.018
	+1	RT-COT	Provide for the use of Standard Commercial of-the-shelf (COTS) Parts	0.010	0.062	0.004
	+1	RT-EQI	Provide for Equipment Interchangability	0.010	0.129	0.009
PICE	-1	IT-MSA	Multi-Site Applicability	0.423	0.153	0.492
	-1	IT-WOO	Possibility of Interference in Implementation (Window of Opportunity)	0.195	0.175	0.259
	-1	IT-FMC	Possibility of Flight Manifest Changes	0.137	0.077	0.080
	-1	IT-MSC	Effects on Multi-System Configuration Systems	0.116	0.032	0.028
	-1	IT-EOH	Possibility of Equipment and Occupational Hazards	0.065	0.029	0.014
	-1	IT-SSR	Site Specific Restrictions	0.033	0.057	0.014
	+1	IT-TCH	Ability to Meet New Technology Considerations	0.031	0.478	0.113

Table-14: Identify Probabilities of Occurrence of Each Subcriterion for all ESRs and Measure the Entropy and Revise the Importance Weight of Stakeholders

Stakeholders	TE K15694	TE K16038	TE K15736/37	TE K15470	TE K15474	TE 0833-56	TE K16055	TE K15998	TE K16032	TE 0835-100	TE K15660	TE K16033	TE FP017-2A	TE 0835-96	TE K16006	TE K15416	TE K16045	TE K16012	TE K15787	TE K15459	TE K16034	TE K16000
	0.90	0.90	0.90	0.90	1.00	1.00	1.00	0.90	0.90	1.00	1.00	0.90	0.90	1.00	1.00	1.00	0.90	1.00	1.00	0.90	1.00	1.00
	0.90	1.00	0.90	0.90	1.00	0.90	1.00	0.90	1.00	1.00	1.00	1.00	0.90	1.00	1.00	0.90	1.00	0.90	0.90	1.00	1.00	1.00
Safety	0.90	0.90	0.90	0.90	0.90	0.90	1.00	0.60	0.60	0.90	1.00	0.60	0.80	0.90	0.90	0.90	0.80	0.90	0.90	0.90	1.00	1.00
	1.00	1.00	1.00	1.00	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.90	1.00	1.00	1.00	1.00	1.00
	1.00	1.00	1.00	1.00	0.60	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.90	1.00	0.80	1.00	1.00	1.00	1.00	1.00
	0.20	0.20	0.10	0.60	0.00	0.00	0.00	0.60	0.00	0.00	0.00	0.00	0.10	0.00	0.00	0.80	0.60	0.40	0.00	0.00	0.00	0.00
	0.20	0.70	0.60	0.60	0.20	0.40	0.20	0.70	0.00	0.30	0.20	0.00	0.30	0.20	0.40	0.80	0.60	0.40	0.40	0.30	0.40	0.00
Systems	0.80	0.30	0.40	0.40	0.80	0.60	0.80	0.30	1.00	0.70	0.80	1.00	0.70	0.80	0.60	0.20	0.40	0.60	0.60	0.70	0.60	1.00
Engineering	0.30	0.70	0.70	0.30	0.20	0.80	0.20	0.60	0.00	0.40	0.00	0.00	0.80	0.00	0.00	0.80	0.60	0.70	0.20	0.50	0.80	0.70
	0.00	0.80	0.80	0.00	0.00	0.90	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.70	0.00	0.00	0.00	0.70
	0.80	0.20	0.80	0.70	0.70	0.30	0.70	0.70	0.30	0.20	0.70	0.30	0.70	0.70	0.20	0.70	0.20	0.70	0.30	0.80	0.20	0.20
	0.60	0.60	0.50	0.60	0.70	0.60	0.60	0.50	0.60	0.60	0.70	0.70	0.60	0.70	0.60	0.60	0.50	0.60	0.60	0.60	0.70	0.60
Program	0.80	0.80	0.70	0.70	0.80	0.80	0.80	0.70	0.80	0.70	0.80	0.80	0.80	0.80	0.80	0.80	0.50	0.70	0.70	0.80	0.80	0.80
Office	0.60	0.60	0.50	0.60	0.60	0.70	0.60	0.60	0.70	0.50	0.70	0.70	0.60	0.70	0.50	0.70	0.70	0.60	0.60	0.60	0.70	0.60
	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.00	0.50	0.50	0.50	0.50	0.50	0.50	0.50
	0.50	0.70	0.70	0.30	0.30	0.60	0.20	0.70	0.20	0.50	0.80	0.20	0.70	0.40	0.00	0.20	0.20	0.30	0.40	0.30	0.30	0.80
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1.00	0.70	0.20	0.00	0.00	0.20	0.00	0.40	1.00	0.00	1.00	1.00	0.60	0.00	1.00	0.20	0.20	0.00	0.40	0.00	1.00	0.40
	0.90	0.70	0.70	0.20	0.00	0.20	0.60	0.40	1.00	0.00	1.00	1.00	0.60	0.00	1.00	0.70	0.70	0.00	0.70	1.00	0.70	0.70
Operations	1.00	0.70	0.70	0.20	0.40	0.20	0.60	0.40	1.00	0.00	1.00	1.00	0.60	0.70	1.00	0.70	0.70	0.00	0.90	1.00	0.70	0.70
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.00	0.70	0.00	0.00	0.00	0.70	0.00	0.70	0.00	0.70	0.70	0.00	0.00	0.00	0.00	1.00	0.00	0.40	0.00	0.00	0.00	0.00
	0.00	0.70	0.70	0.00	0.00	0.70	0.00	0.00	0.70	0.00	0.00	0.70	0.00	0.00	0.00	0.70	0.00	0.00	0.00	0.00	0.70	0.70
	0.00	0.70	0.70	0.00	0.00	0.70	0.00	0.00	0.00	0.00	0.70	0.00	0.00	0.70	0.00	0.70	0.00	0.00	0.00	0.70	0.70	0.70
Reliability	0.00	0.70	0.70	0.00	0.00	0.70	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.70	0.00	0.00	0.00	0.00	0.00	0.00
	0.00	0.70	0.00	0.00	0.00	0.70	0.00	0.00	0.70	0.00	0.00	0.70	0.00	0.00	0.00	0.70	0.00	0.00	0.00	0.00	0.00	0.70
	0.00	0.70	0.70	0.00	0.00	0.70	0.00	1.00	0.00	0.00	0.00	0.00	0.70	0.00	0.00	0.70	0.00	0.70	0.00	0.00	0.00	0.70
	0.00	0.70	0.70	0.00	0.00	0.70	0.00	0.00	0.70	0.00	0.00	0.70	0.00	0.00	0.00	0.70	0.00	0.00	0.00	0.00	0.00	0.70
	0.70	0.70	0.70	0.00	0.00	0.70	0.00	0.00	0.70	0.00	0.70	0.70	0.00	0.70	0.00	0.70	0.00	0.00	0.00	0.70	0.00	0.70
	0.70	0.70	0.70	0.00	0.00	0.70	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.60	0.20	0.30	0.80	0.80	0.50	0.80	0.40	0.90	0.60	1.00	0.80	0.50	0.50	0.50	0.10	0.50	0.90	0.50	0.50	0.90	0.90
	0.60	0.20	0.60	0.40	0.10	0.20	0.00	0.80	0.80	0.50	0.00	0.80	0.30	0.20	0.30	0.60	0.20	0.80	0.20	0.00	0.80	0.80
	0.50	0.30	0.70	0.30	0.20	0.10	0.50	0.80	0.80	0.80	0.60	0.80	0.50	0.30	0.40	0.50	0.20	0.70	0.70	0.40	0.80	0.70
PICB	0.70	0.40	0.90	0.80	0.50	0.30	0.90	0.90	0.90	0.80	0.70	0.90	0.90	0.30	0.90	0.70	0.80	0.60	0.90	0.30	0.90	0.70
	0.70	0.80	0.80	0.80	0.80	0.90	0.60	0.80	0.80	0.50	0.70	0.80	0.60	0.70	0.80	0.50	0.90	0.80	0.60	0.40	0.90	0.80
	0.10	0.40	0.60	0.60	0.30	0.50	0.20	0.90	0.80	0.50	0.80	0.70	0.80	0.40	0.70	0.80	0.50	0.60	0.80	0.80	0.40	0.80
	0.00	0.10	0.00	0.20	0.00	0.20	0.00	0.10	0.00	0.10	0.10	0.00	0.00	0.10	0.10	0.00	0.00	0.10	0.10	0.10	0.10	0.10

Table-15: Identify Probabilities of Occurrence of Each Subcriterion for all ESRs and Measure the Entropy and Revise the Importance Weight of Stakeholders

Stakeholders	TV K15747	TV K16026	TV K15859	TV K14997	TV K15445	TV K15995	TV K14515	TV K15512	TV K15497	TV K15936	TV K15604	TV K15442	TV K15504	TV K15535	TV K15770	TV K15943	TV K14920	TV K15542	TV K15369
	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.90	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1.00	0.90	0.90	1.00	1.00	1.00	1.00	0.90	1.00	0.90	0.90	0.90	0.90	0.90	1.00	1.00	1.00	1.00	1.00
Safety	0.60	0.90	1.00	1.00	1.00	0.90	1.00	0.90	0.80	0.70	0.70	0.70	0.70	0.90	1.00	0.90	1.00	0.90	0.90
	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.60	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.10	0.00	0.10	0.00	0.00	0.20	0.50	0.00	0.00	0.00	0.00	0.40	0.20	0.00	0.00	0.10	0.10	1.00	0.40
	0.10	0.60	0.70	0.80	0.70	0.70	0.50	0.00	0.00	0.00	1.00	0.00	0.00	0.70	0.00	0.00	0.00	0.00	1.00
Systems Engineering	0.30	1.00	0.30	0.10	0.00	0.20	0.40	1.00	1.00	1.00	0.00	0.40	0.80	1.00	1.00	0.30	0.30	1.00	1.00
	0.00	0.00	0.70	1.00	0.00	0.50	0.90	0.00	0.00	0.00	0.00	0.80	0.00	0.70	0.00	0.00	0.90	0.10	1.00
	0.00	0.00	0.10	0.10	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
	0.70	0.50	0.70	0.30	0.30	0.30	0.30	0.60	0.70	0.30	0.70	0.50	0.50	0.30	0.20	0.50	0.40	0.60	0.40
	0.50	0.60	0.60	0.60	0.60	0.60	0.60	0.60	0.60	0.60	0.60	0.60	0.60	0.50	0.70	0.60	0.60	0.60	0.60
Program Office	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80
	0.70	0.60	0.60	0.70	0.60	0.50	0.70	0.70	0.70	0.70	0.70	0.60	0.70	0.70	0.70	0.70	0.70	0.70	0.70
	0.50	0.50	0.50	0.50	0.50	0.50	0.60	0.50	0.50	0.50	0.60	0.50	0.50	0.50	0.60	0.50	0.50	0.50	0.50
	0.30	0.50	0.70	0.70	0.60	0.20	0.80	0.10	0.10	0.10	0.20	0.10	0.50	0.40	0.30	0.60	0.60	0.20	0.20
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Operations																			
	0.30	0.60	0.20	0.60	0.70	0.20	0.20	0.20	0.60	0.40	0.20	0.20	0.20	1.00	0.20	0.20	0.40	0.10	0.10
	0.60	0.50	0.80	1.00	0.20	0.70	0.30	0.90	0.20	0.60	0.60	0.80	0.40	1.00	0.20	0.20	0.60	0.30	0.10
	1.00	0.20	0.10	0.10	0.30	0.10	0.10	0.80	1.00	0.80	0.10	0.10	0.10	0.10	0.10	0.30	0.10	0.20	0.10
	0.10	0.10	0.10	0.40	1.00	0.10	1.00	0.10	0.40	0.10	0.40	0.10	0.10	1.00	0.10	0.10	0.20	0.10	0.10
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
	0.00	0.00	0.70	0.70	0.70	0.00	1.00	0.00	0.00	0.70	0.00	0.00	0.70	0.00	0.00	0.00	0.00	0.00	0.00
	0.00	0.00	0.00	0.70	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.70	0.00	0.00	1.00	0.00	0.00	0.70
Reliability	0.00	0.00	0.00	0.70	0.70	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.70	0.00	0.00	0.00	0.00	0.00	0.00
	0.00	0.00	0.00	0.70	0.70	0.00	0.00	0.00	0.00	0.70	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.00	0.00	0.70	0.70	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.70	0.70	0.00	0.00	0.00	0.00	0.00	0.00
	0.00	0.00	0.00	0.70	0.70	0.00	1.00	0.00	0.00	0.70	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.00	0.00	0.00	0.70	0.70	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.70	0.00	0.00	0.00	0.00	0.00	0.70
	0.00	0.00	0.00	0.70	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.00	0.00	0.00	0.80	0.50	0.50	0.90	0.90	0.90	0.50	0.00	0.20	0.80	0.80	0.80	0.60	0.80	0.80	0.40
	0.00	0.90	1.00	0.90	0.50	0.20	0.70	0.50	0.90	0.40	0.90	0.20	0.90	0.80	0.60	0.80	0.90	0.80	0.30
	0.30	0.80	0.90	0.90	0.50	0.20	0.70	0.90	0.90	0.40	0.80	0.20	0.80	0.80	0.70	0.40	0.90	0.90	0.50
PICB	0.50	0.90	0.90	0.80	0.70	0.60	0.80	0.90	1.00	0.80	1.00	0.60	0.60	0.60	0.80	0.80	0.70	0.70	0.60
	0.20	1.00	1.00	0.50	0.40	0.70	0.60	0.40	0.80	0.50	1.00	0.50	0.80	0.70	0.80	0.70	0.80	0.70	0.80
	0.10	0.90	0.90	0.80	0.50	0.30	0.70	0.90	0.80	0.50	1.00	0.30	0.70	0.80	0.60	0.60	0.60	0.70	0.90
	0.60	0.00	0.00	0.30	0.10	0.10	0.10	0.30	0.00	0.10	0.00	0.00	0.10	0.40	0.10	0.10	0.40	0.10	0.60

Tabale-16: Identify the Individual Project Success Factor of each ESR and Develop Rankings and Develop Consensus Ranking of ESRs

ESR NO.	Organization	GSWT Voting Committee Rankings					Consensus Rankings
		M. Allison	D. Kelley	S. Lee	B. Tootill	J. White	
K15616	TE	1	1	1	1	1	1
K15562	TV	2	2	2	2	2	2
K14515	TV	3	3	6	4	3	3
K15660	TE	8	6	3	3	8	4
K16012	TE	7	4	5	7	5	5
K15998	TE	4	5	4	9	4	6
K15569	TV	5	7	14	10	6	7
K14887	TV	13	8	9	5	11	8
K15642	TV	12	10	12	12	10	9
K16038	TE	6	18	15	8	7	10
K16036	TE	11	11	18	6	15	11
K15535	TV	9	12	20	11	14	12
K15694	TE	15	16	7	17	13	13
K15470	TE	14	9	11	23	9	14
K15645	TV	16	15	19	13	16	15
K15858	TV	18	13	10	18	17	16
K16000	TE	17	14	21	14	19	17
GS33-56	TE	19	21	16	15	20	18
K15736/37	TE	21	22	8	22	18	19
K15504	TV	24	20	17	19	21	20
K14920	TV	22	17	23	20	22	21
K16045	TE	10	23	33	25	12	22
K15985	TV	23	24	32	24	24	23
FF017-2A	TE	26	31	22	30	25	24
K16006	TE	20	27	36	21	26	25
K15512	TV	30	26	26	26	30	26
K16026	TV	29	28	31	27	29	27
K16055	TE	34	25	25	31	33	28
K15684	TV	32	32	27	34	31	29
K16032	TE	27	37	34	28	27	30
GS35-86	TE	36	29	24	32	35	31
K16033	TE	28	39	35	29	28	32
K15787	TE	31	33	38	33	34	33
K15697	TV	33	35	29	37	32	34
K15674	TE	38	30	28	36	36	35
GS35-100	TE	35	34	39	35	37	36
K15563	TV	25	19	13	16	23	37
K15659	TE	40	38	30	39	38	38
K15770	TV	37	36	40	38	40	39
K15747	TV	39	40	37	40	39	40
K15936	TV	41	41	41	41	41	41

Table-17: Ranking of Tactical ESRs Using Consensus Ranking Approach

Rank	ESR NO.	Organization	Category	ROM	Cumulative ROM	PSF
1	K15616	TE	2E	46,600	46,600	0.803
2	K15562	TV	2E	34,830	81,430	0.656
3	K14515	TV	2M	117,647	199,077	0.619
4	K15660	TE	2E	15,000	214,077	0.637
5	K16012	TE	2S	40,000	254,077	0.605
6	K15998	TE	2S	112,000	366,077	0.616
7	K15569	TV	2M	61,570	427,647	0.582
8	K14887	TV	2M	20,680	448,327	0.602
9	K15642	TV	2S	17,077	465,404	0.576
10	K16038	TE	2M	56,600	522,004	0.590
11	K16036	TE	2E	14,200	536,204	0.599
12	K15535	TV	2E	110,662	646,866	0.592
13	K15694	TE	2E	5,700	652,566	0.601
14	K15470	TE	2S	82,263	734,829	0.570
15	K15645	TV	2O	26,942	761,771	0.576
16	K15858	TV	2O	55,000	816,771	0.573
17	K16000	TE	2E	35,000	851,771	0.575
18	GS33-56	TE	2M	50,000	901,771	0.563
19	K15736/37	TE	2S	175,060	1,076,831	0.569
20	K15504	TV	2S	64,383	1,141,214	0.563
21	K14920	TV	2M	111,500	1,252,714	0.563
22	K16045	TE	2S	20,200	1,272,914	0.544
23	K15985	TV	2E	49,200	1,322,114	0.537
24	FF017-2A	TE	2M	50,000	1,372,114	0.552
25	K16006	TE	2E	14,400	1,386,514	0.555
26	K15512	TV	2E	19,316	1,405,830	0.544
27	K16026	TV	2E	8,000	1,413,830	0.538
28	K16055	TE	2E	29,000	1,442,830	0.531
29	K15684	TV	2S	13,785	1,456,615	0.530
30	K16032	TE	2S	16,000	1,472,615	0.550
31	GS35-86	TE	2M	7,200	1,479,815	0.524
32	K16033	TE	2S	22,000	1,501,815	0.548
33	K15787	TE	2M	46,000	1,547,815	0.521
34	K15697	TV	2S	20,196	1,568,011	0.540
35	K15674	TE	2S	18,000	1,586,011	0.521
36	GS35-100	TE	2E	70,000	1,656,011	0.507
37	K15563	TV	2M	1,850	1,657,861	0.568
38	K15659	TE	2S	30,000	1,687,861	0.513
39	K15770	TV	2M	2,000	1,689,861	0.505
40	K15747	TV	2S	240,029	1,929,890	0.502
41	K15936	TV	2S	29,000	1,958,890	0.487
			Overall Mean			0.567

Table-18: Ranking of ESRs with Safety Impact (2S)

Rank	ESR NO.	Organization	Category	ROM	CUM ROM	PSF
5	K16012	TE	2S	40,000	40,000	0.605
6	K15998	TE	2S	112,000	152,000	0.616
9	K15642	TV	2S	17,077	169,077	0.576
14	K15470	TE	2S	82,263	251,340	0.570
19	K15736/37	TE	2S	175,060	426,400	0.569
20	K15504	TV	2S	64,383	490,783	0.563
22	K16045	TE	2S	20,200	510,983	0.544
29	K15684	TV	2S	13,785	524,768	0.530
30	K16032	TE	2S	16,000	540,768	0.550
32	K16033	TE	2S	22,000	562,768	0.548
34	K15697	TV	2S	20,196	582,964	0.540
35	K15674	TE	2S	18,000	600,964	0.521
38	K15659	TE	2S	30,000	630,964	0.513
40	K15747	TV	2S	240,029	870,993	0.502
41	K15936	TV	2S	29,000	899,993	0.487
				Mean		**0.549**

Tble-19: Ranking of ESRs Related to Obsolescence Issues (2O)

Rank	ESR NO.	Organization	Category	ROM	CUM ROM	PSF
15	K15645	TV	2O	26,942	26,942	0.576
16	K15858	TV	2O	55,000	81,942	0.573
				Mean		**0.575**

Table-20: Ranking of ESRs Related to High Maintenance/Design Issues (2M)

ESR NO.	Organization	Category	ROM	CUM ROM	PSF
K14515	TV	2M	117,647	117,647	0.619
K15569	TV	2M	61,570	179,217	0.582
K14887	TV	2M	20,680	199,897	0.602
K16038	TE	2M	56,600	256,497	0.590
GS33-56	TE	2M	50,000	306,497	0.563
K14920	TV	2M	111,500	417,997	0.563
FF017-2A	TE	2M	50,000	467,997	0.552
GS35-86	TE	2M	7,200	475,197	0.524
K15787	TE	2M	46,000	521,197	0.521
K15563	TV	2M	1,850	523,047	0.568
K15770	TV	2M	2,000	525,047	0.505
			Mean		**0.563**

Table-21: Ranking of ESRs Related to Enhancement Issues (2E)

Rank	ESR NO.	Organization	Category	ROM	CUM ROM	PSF
1	K15616	TE	2E	46,600	46,600	0.803
2	K15562	TV	2E	34,830	81,430	0.656
4	K15660	TE	2E	15,000	96,430	0.637
11	K16036	TE	2E	14,200	110,630	0.599
12	K15535	TV	2E	110,662	221,292	0.592
13	K15694	TE	2E	5,700	226,992	0.601
17	K16000	TE	2E	35,000	261,992	0.575
23	K15985	TV	2E	49,200	311,192	0.537
25	K16006	TE	2E	14,400	325,592	0.555
26	K15512	TV	2E	19,316	344,908	0.544
27	K16026	TV	2E	8,000	352,908	0.538
28	K16055	TE	2E	29,000	381,908	0.531
36	GS35-100	TE	2E	70,000	451,908	0.507
				Mean		**0.590**

Future Tasks

- **Develop a template for ESR originators to provide more complete and consistent information about ESRs to the GSWT steering committee and stakeholders.**

- **Develop an objective and structured process for the initial screening of ESRs.**

- **Investigate the possibility of an organization-wide implementation of the model.**

- **Extensive training of the stakeholders and the GSWT steering committee members to eliminate the need for a facilitator.**

- **Automate the process as much as possible.**

- **Revisit the Structure of the GSWT steering committee.**

- **Create an environment to compile PSF data on ESRs and move towards a disciplined system that could be used to address supportbility threshold issues.**

1995 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

5/9-47
7759
p.32

LIGHTNING ELECTROMAGNETICS

Dr. Parveen Wahid
Associate Professor
Department of Electrical and Computer Engineering
University of Central Florida
Orlando, Florida

KSC Colleague - Carl Lennon
Atmospheric Science

Contract Number NASA-NGT-60002
Supplement 19

August 14, 1995

PAGE 536 INTENTIONALLY BLANK

ACKNOWLEDGEMENTS

The experience of participating in the NASA/ASEE Summer Faculty Fellowship Program was exciting and invigorating. Special thanks are extended to Mr. Carl Lennon, my NASA colleague, for his tremendous help and support in all aspects of the project. His great enthusiasm and wealth of knowledge in the field of lightning research contributed to a large extent in making the summer research work interesting and exciting. The author would like to thank Launa Maier for providing all the data necessary to conduct the research and for helping out on various occasions in spite of her own busy schedule. John Kiriazes helped provide access to the Internet and Steve Schaefer provided useful information on the calibration of the antennas.

Thanks are due to Dr. Ray Hosler, Director of the NASA/ASEE Summer Faculty Fellowship Program and to Ms. Kari Stiles for putting together a well coordinated program. The tours at NASA and the presentations by NASA personnel were interesting and informative.

ABSTRACT

This project involved the determination of the effective radiated power of lightning sources and the polarization of the radiating source. This requires the computation of the antenna patterns at all the LDAR site receiving antennas. The known radiation patterns and RF signal levels measured at the antennas will be used to determine the effective radiated power of the lightning source. The azimuth and elevation patterns of the antennas in the LDAR system were computed using flight test data that was gathered specifically for this purpose. The results presented in this report deal with the azimuth patterns for all the antennas and the elevation patterns for three of the seven sites.

SUMMARY

The objective of the research was to determine the effective radiated power of lightning sources and to arrive at a model for the antenna pattern and polarization for a lightning produced RF source . This data is useful in understanding the development of thunderstorms and for possibly predicting the end of a storm. In order to achieve this goal, flight test data and lightning data were collected at the seven antenna sites of the lightning detection and ranging (LDAR) system at Kennedy Space Center. Flight test data was gathered by flying an aircraft across each of the sites and also in circles around each site. The aircraft was equipped with a horizontal antenna at its tail and a vertical antenna at the bottom which allowed data on vertical and horizontal polarization to be obtained. From this data it is possible to compute the azimuth and elevation patterns of the antennas.

The LDAR system consists of seven antennas operating at a frequency of 66 MHZ. These antennas are conical spiral antennas providing broadband hemispherical coverage. The antennas record the signal levels from lightning occurrences at each antenna site. Using this information, the antenna patterns and the location of the antennas it is possible to model the antenna pattern of the lightning source. Information regarding the polarization of the source can also be deduced from this antenna pattern.

The research work conducted under this project was mainly theoretical. Vast amounts of lightning data and flight test data were available to carry out the necessary computations. During the time that was available the first part of this effort was completed. The azimuth patterns for all the antennas were computed for both horizontal and vertical polarizations. These patterns will have to be analyzed to look for any possible effects of structures close to the antenna site. The elevation patterns were determined for three of the seven antennas also for horizontal and vertical polarizations. Manufacturer's data on the antenna patterns are available for circular polarization. Due to the limited amount of time the lightning data could not be analyzed.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

I. INTRODUCTION

The Lightning Detection and Ranging (LDAR) system was implemented at the Kennedy Space Center (KSC) in 1991. The Kennedy Space Center is located in a region where there are frequent lightning strikes throughout the year. Lightning activity or even the threat of lightning affects routine operations at KSC, the shuttle launches and associated operations. The LDAR system was developed in order to provide accurate information concerning lightning activity and potential lightning hazard in the KSC area.

1.1 THE LDAR SYSTEM

The LDAR system maps the location of cloud-to-ground lightning and in-cloud lightning at a frequency of 66 MHZ. It consists of one central site and 6 remote sites. The antennas at these sites detect lightning induced disturbances at this frequency and using the time of arrival of the radiation at the different sites a three dimensional position of the lightning source is computed [1,2]. The LDAR system generates up to 10,000 data points per second for a lightning flash and provides projections of the lightning in near-real time varying from 1 second to 2 minutes.

The LDAR antennas are located within a radius of 6 to 10 km from the LDAR central site. All the antennas are conical spiral left circularly polarized antennas. Lightning-produced pulses received st the remote sites are processed by log video detectors and the video pulses are transmitted to the central site via microwave links. The physical configuration of the system is such that the direct RF pulses arrive at the central site prior to the retransmitted pulses from the remote sites. The system is triggered when the pulse exceeds a certain threshold at the central site and a 100 microseconds data analysis period commences. During this analysis period the system determines the time of occurrence of the largest amplitude pulse for each of the sites. At the end of the analysis period, the time and amplitude data for each site is collected and tagged with the time of the day to the nearest microsecond. The system is rearmed within 10 microseconds. After the data has been gathered, it is transmitted to a set of computers for testing, calculations of the source locations and display.

The results presented in this report deal with the calculation of the antenna patterns for the LDAR system. The azimuth and elevation patterns computed using flight test data are presented.

1. 2 LDAR SITE INFORMATION

Survey data for the LDAR antenna sites is available as the WGS-84 geodetic latitude, longitude and the mean sea elevation of the antennas. Using the appropriate coordinate transformations the x, y, z locations of the sites are calculated with the central site being at (0,0,0).

The geodetic height for the central site used is the value measured to the top of the antenna. The reference geoid for the GPS (WGS-84) is 28.19 m . Figure 1-1 shows the locations of the LDAR sites and Table 1 provides the x,y,z coordinates for the sites.

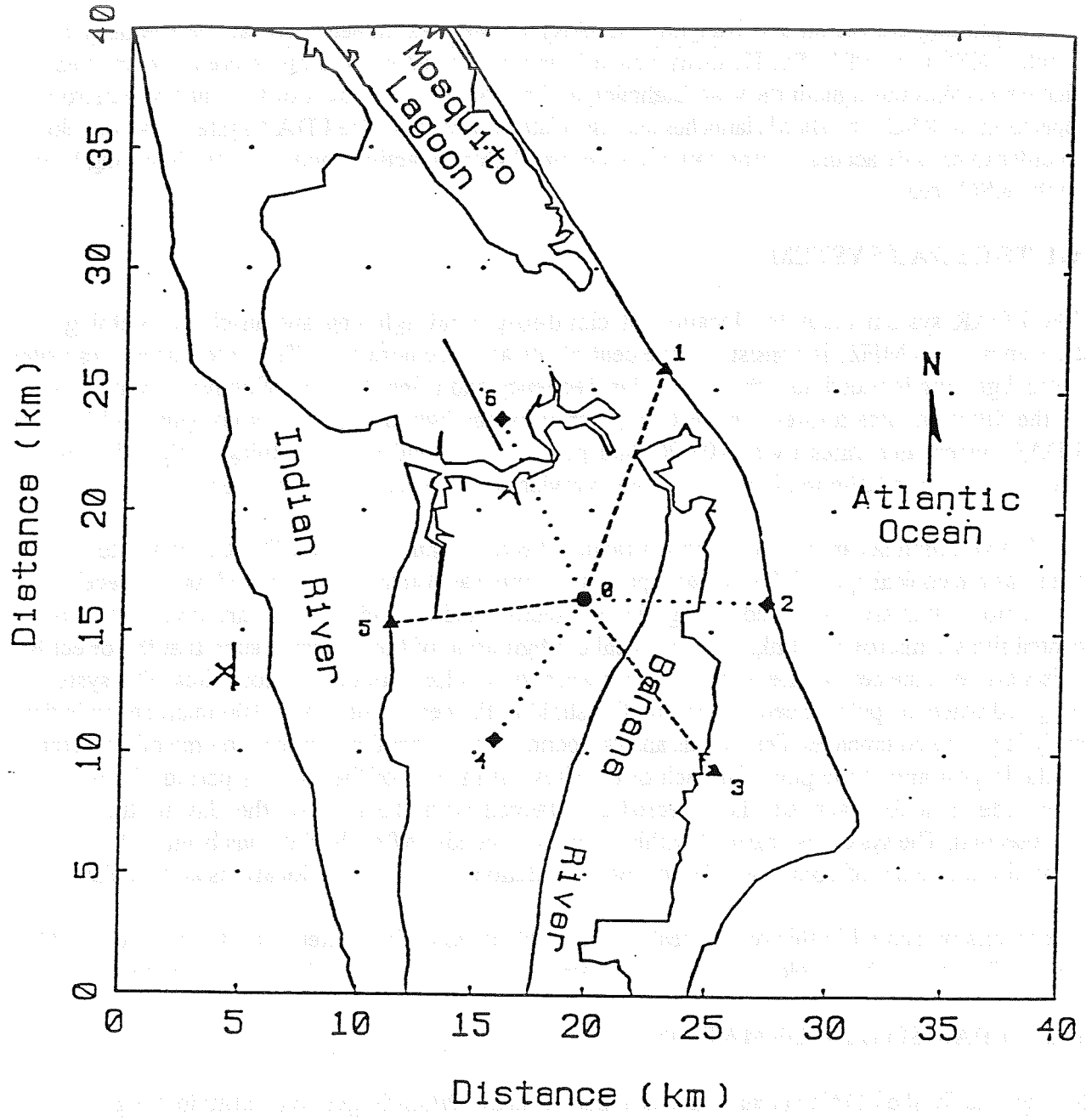


Figure 1-1: LDAR site locations

Table 1: Geodetic data and the x, y, z coordinates of the LDAR antennas

Site #	Latitude (d, m, s)	Longitude (d, m, s)	Height (m)	Mean Sea Level Elevation (m)
0	28 32 19.47637	80 38 34.29894	-20.00580	4.2670
1	28 37 26.24000	80 36 35.10000	-14.25760	13.9324
2	28 32 12.88000	80 33 54.47000	-18.04630	10.1437
3	28 28 29.19000	80 35 11.75000	-16.54970	11.6403
4	28 29 10.09000	80 40 56.06000	- 8.83011	19.3599
5	28 31 45.65000	80 43 44.98000	-11.00940	17.1806
6	28 36 20.79000	80 40 52.73000	-17.38790	10.8021

Site #	X (m)	Y (m)	Z (m)
0	0.97	0.00	0.00
1	3241.46	9443.48	13.59
2	7608.28	-205.54	6.49
3	5507.38	-7090.74	9.79
4	-3852.88	-5830.96	15.02
5	-8445.06	-1044.40	14.67
6	-3762.35	7428.39	8.07

II. FLIGHT TEST DATA

The LDAR flight test was performed on June 1, 1995. The aircraft platform used was a NASA Wallops Island C-I30. A high power pulse amplifier was installed on the aircraft to simulate lightning discharges. The pulse was at 66 MHZ frequency with a pulse repetition frequency of 25 Hz. The position of the aircraft was recorded using a GPS receiver capable of measuring the aircraft position to one meter, rms. The GPS antenna is located on the top of the aircraft in front of the propeller line.

Two GPS systems designated receiver 1 and receiver 2 were used to simultaneously record the aircraft position from the same GPS antenna. The data was recorded as output from the GPS receiver. The receiver outputs an ASCII position record once per second. This record contains the aircraft position in WGS-84 coordinates and earth centered earth fixed coordinates and includes a GPS time tag for all position measurements.

2.1 TEST PLAN

The LDAR antennas are conical spiral left circularly polarized antennas. These are broadband antennas with a frequency range of 60 MHZ to 300 MHZ and provide hemispherical coverage. By using a circularly polarized antenna the variation in signal amplitude due to variation in the polarization of the radiating source is reduced.

Two coaxial antennas, one vertically polarized and the other horizontally polarized, were installed on the aircraft at the bottom and on the tail respectively. These antennas operate at 66 MHZ. The purpose of the flight test is to measure the antenna patterns of the LDAR antennas. The aircraft maintained a flight altitude of 25,000 feet above ground level. The test plan consists of two parts

(i) Fly circles of radius 8.2 statute miles around the 7 LDAR sites. Each circle is flown twice, for the bottom antenna and the tail antenna.

(ii) Fly a clover leaf pattern with each leg extending 8 statute miles beyond the closest LDAR site. The radials are to and from the central site, passing over each remote site. This pattern is also repeated for the bottom and tail antenna.

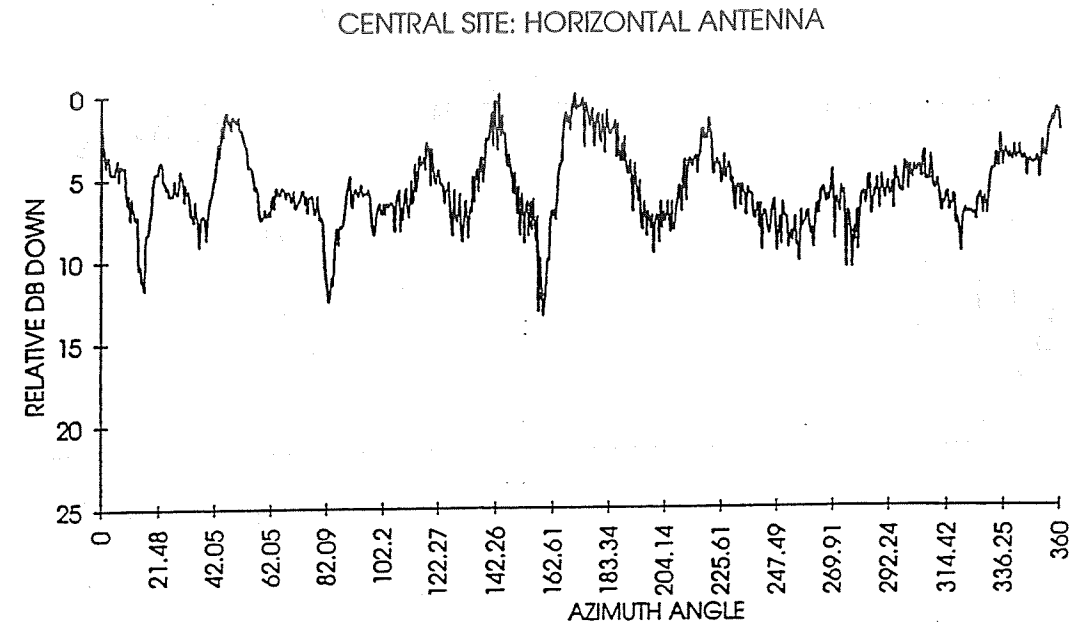
These two flight plans allow the calculation of the azimuth and elevation pattern variation of the antennas. The data recorded provides the x, y, z coordinates of the aircraft and the amplitude levels at all the antenna sites and the UTC time for all position measurements.

The x, y, z positions of the aircraft and its corresponding azimuth and elevation angles calculated with the GPS data were compared with those obtained with the LDAR data. The values for the angles were found to agree to within a tenth of a degree. Only results based on the LDAR data are presented in this report.

All the antennas are calibrated using a high powered impulse generator capable of producing a 400 V peak-to-peak pulse in a 50 ohm load. The calibration data is available for all the sites and is used in calculating the amplitude level of the received signal at the sites.

III. AZIMUTHAL PATTERNS OF THE ANTENNAS

The azimuthal patterns of the antennas at all the seven sites computed using data obtained from the aircraft flying circles around each site are shown in Figures 3-1 to 3-7. In each case the pattern is presented for the vertical and the horizontal antenna. The patterns were obtained for a 30 degree elevation angle. The sharp null (about 20 db down) are due to problems with the transmitter which caused the signal level to drop drastically at all the sites. The azimuthal pattern for these antennas for a circular polarized signal is expected to have very slight variation of the order of 1 to 3 db over the complete range of the azimuth angle. The calculated patterns show a variation of the order of 10 to 15 db in some instances. The antenna at site 1 shows a variation with azimuth angle for both the horizontal antenna and the vertical antenna of about 8 db and appears to have a more acceptable pattern as compared to the other antennas. Sites 4 and 5 show major fluctuations in the amplitude level with angular variation. The antenna patterns are of course influenced by their locations and by structures, buildings etc. in their vicinity. Hence each of these patterns has to be analyzed individually with respect to its location to characterize the effect of its surrounding on the pattern.



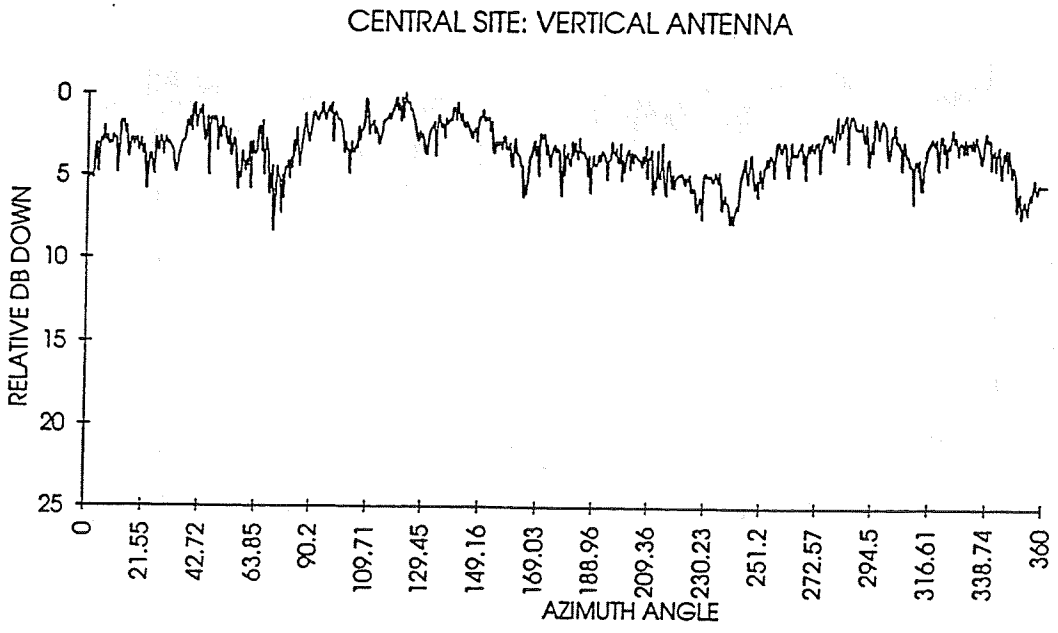
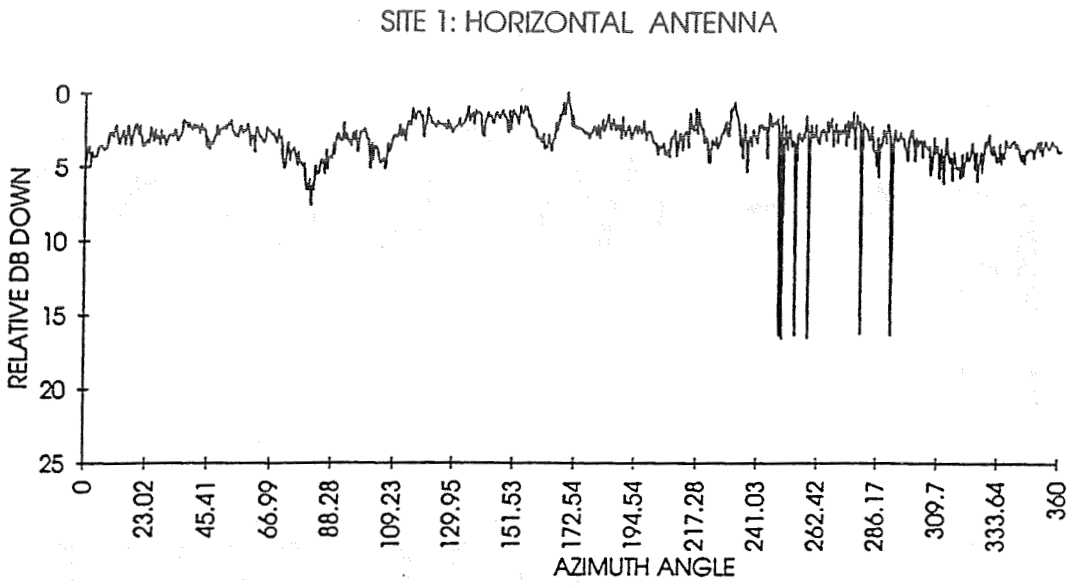


Figure 3-1: Central site azimuthal pattern. (30 degree elevation angle)



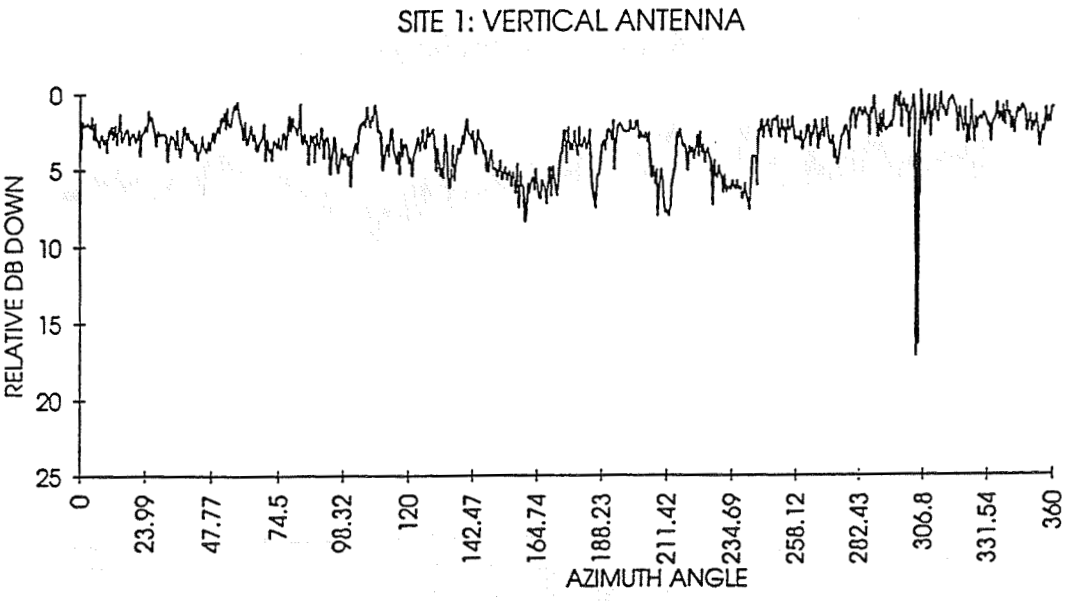
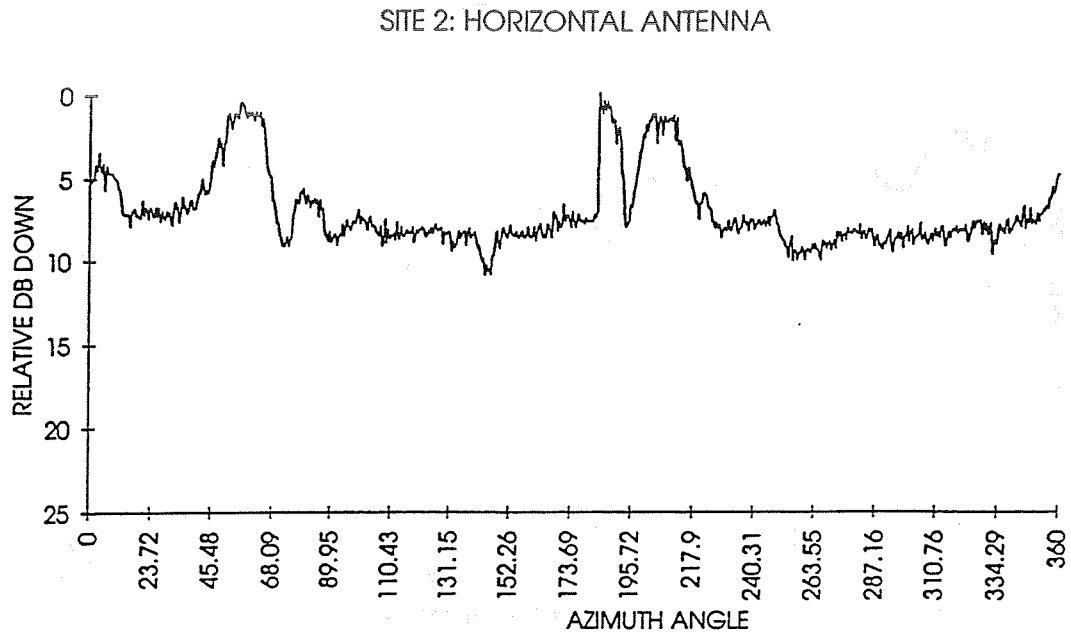


Figure 3-2: Site 1 azimuthal pattern. (30 degree elevation angle)



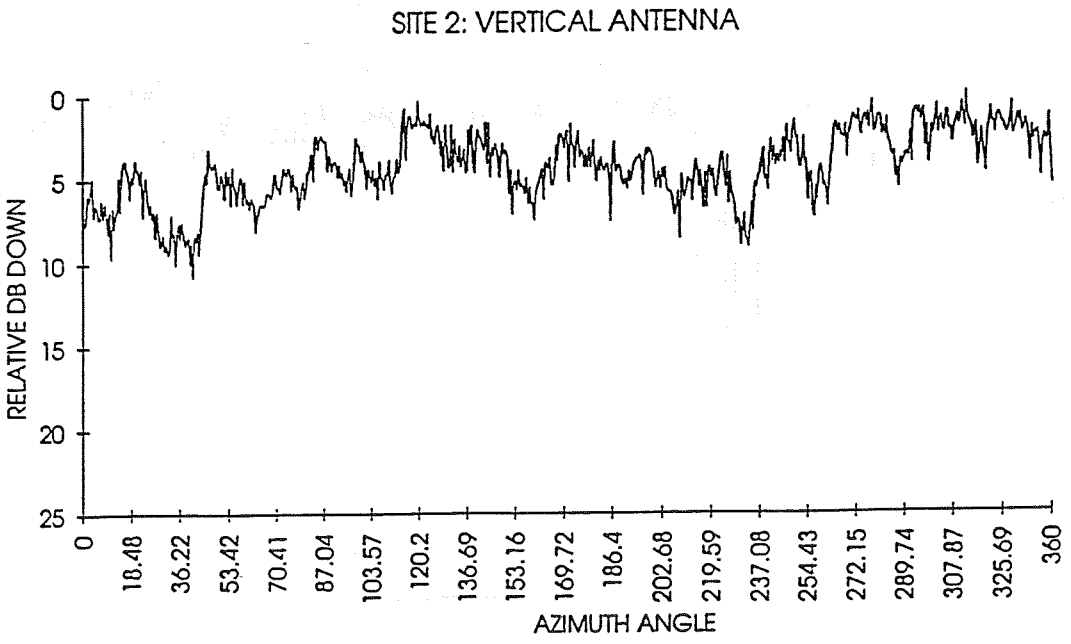


Figure 3-3: Site 2 azimuthal pattern. (30 degree elevation angle)

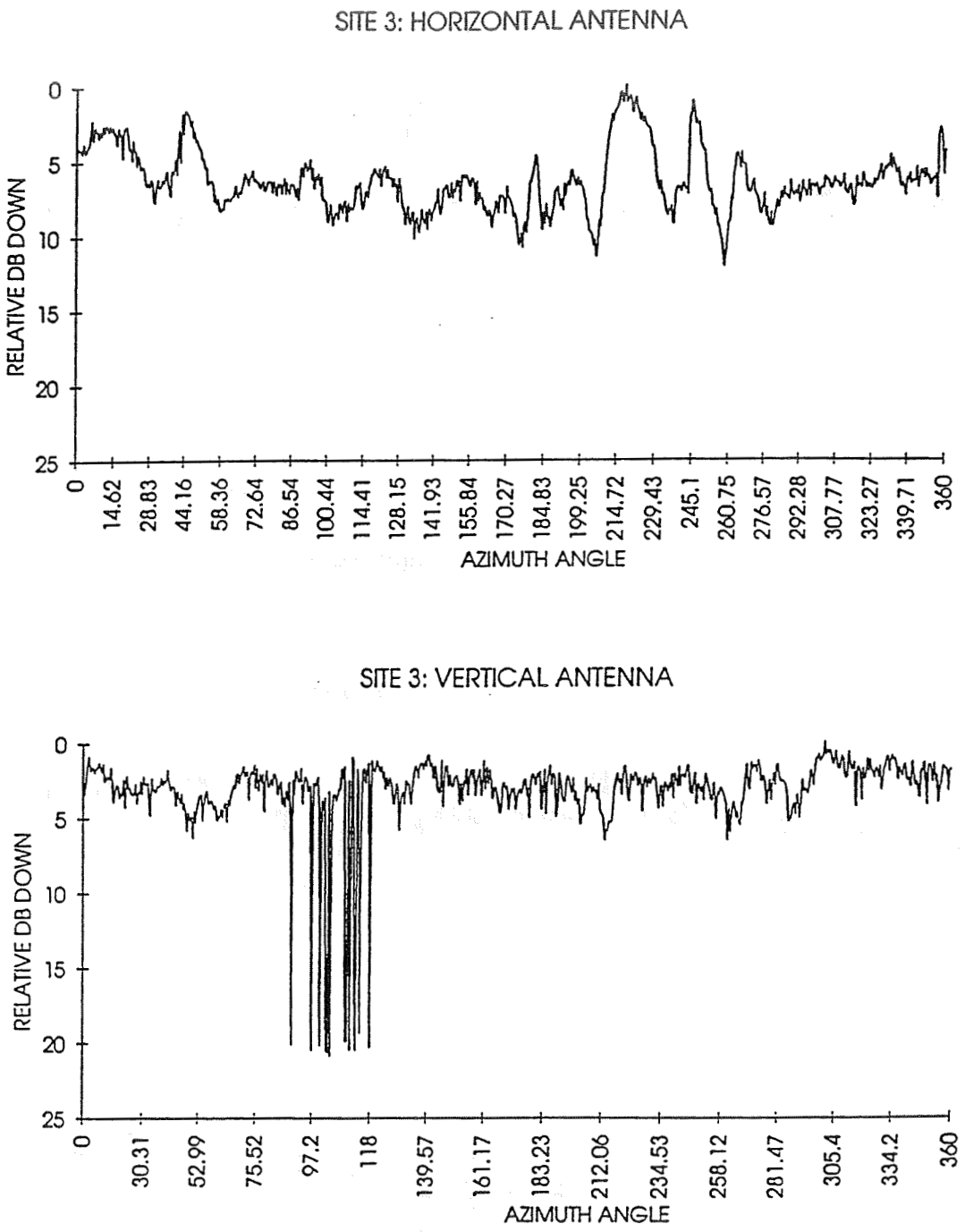


Figure 3-4: Site 3 azimuthal pattern. (30 degree elevation angle)

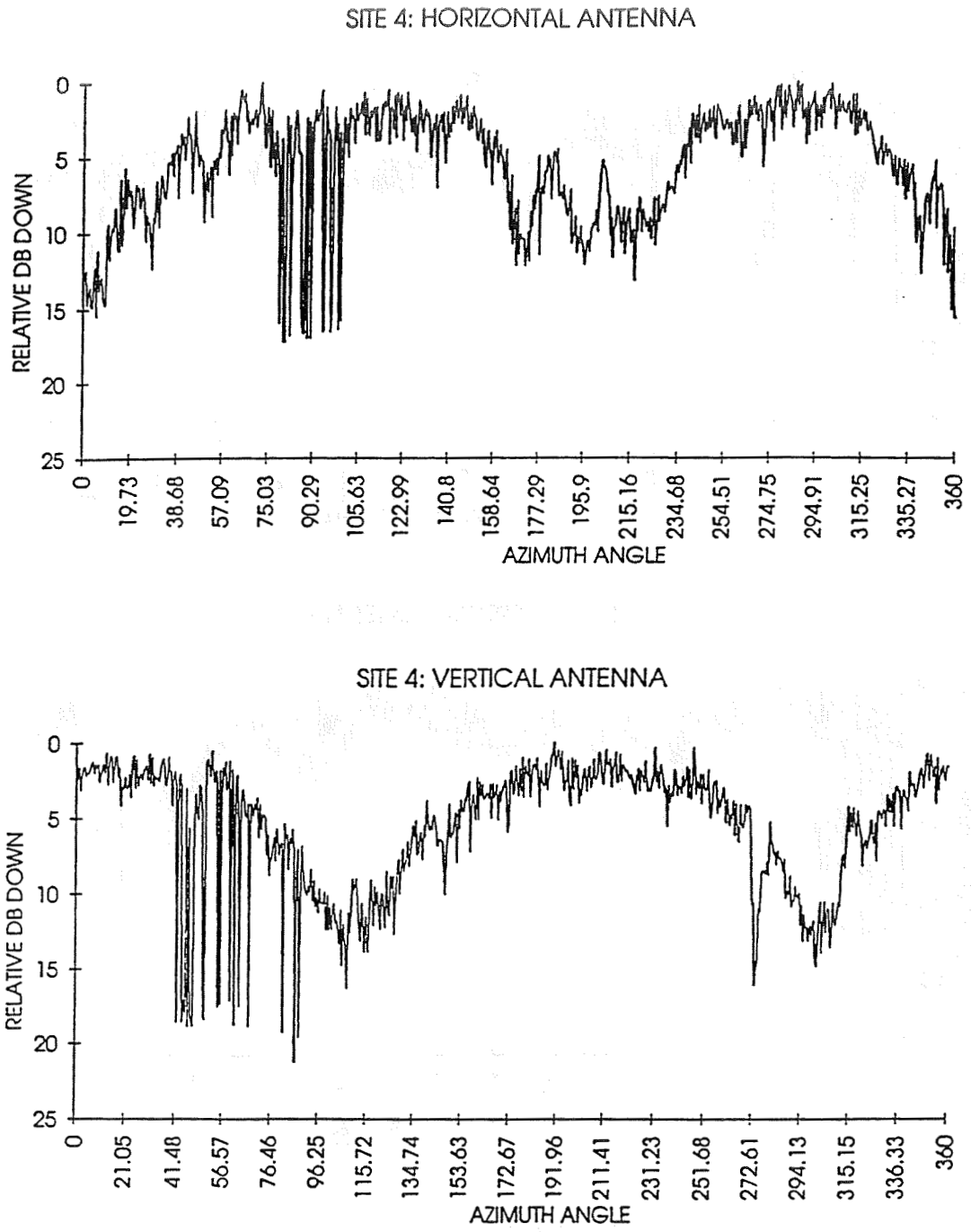
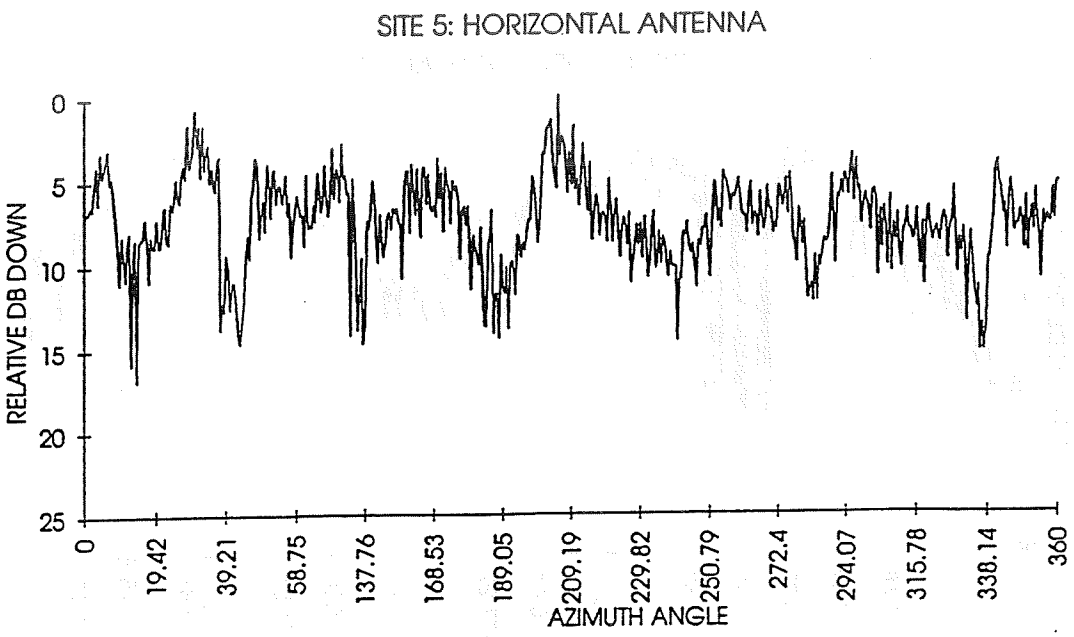


Figure 3-5: Site 4 azimuthal pattern. (30 degree elevation angle)



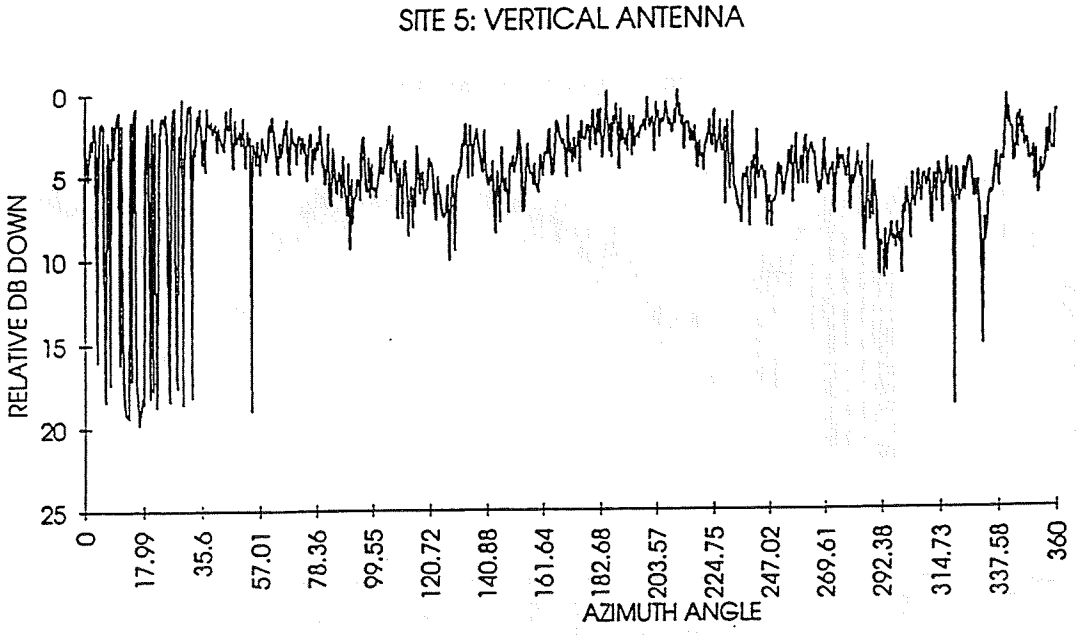
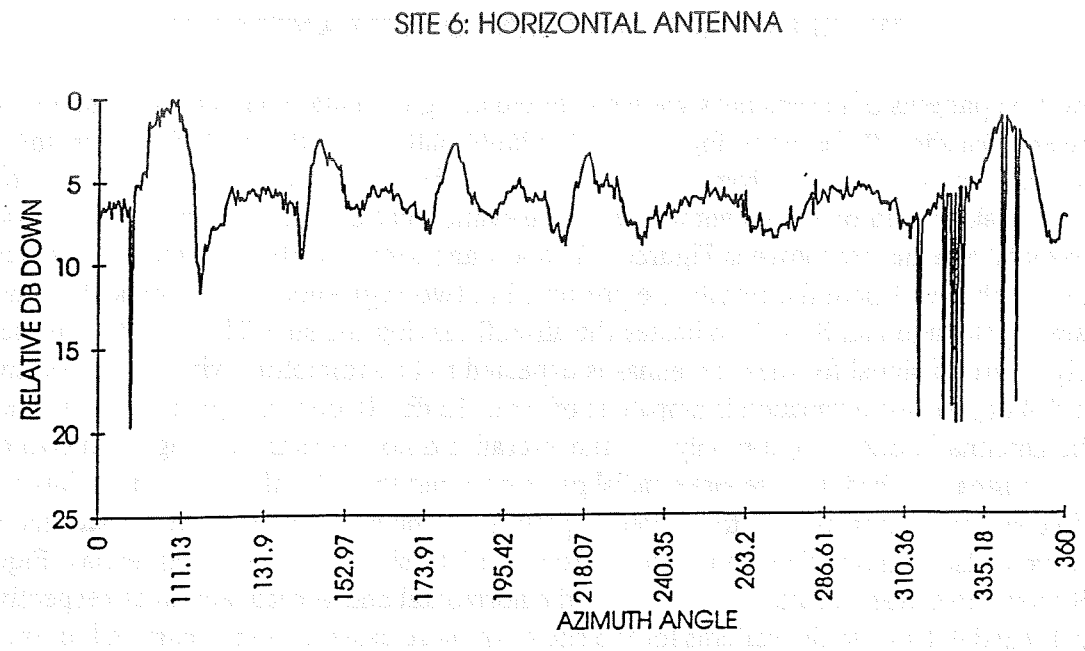


Figure 3-6: Site 5 azimuthal pattern. (30 degree elevation angle)



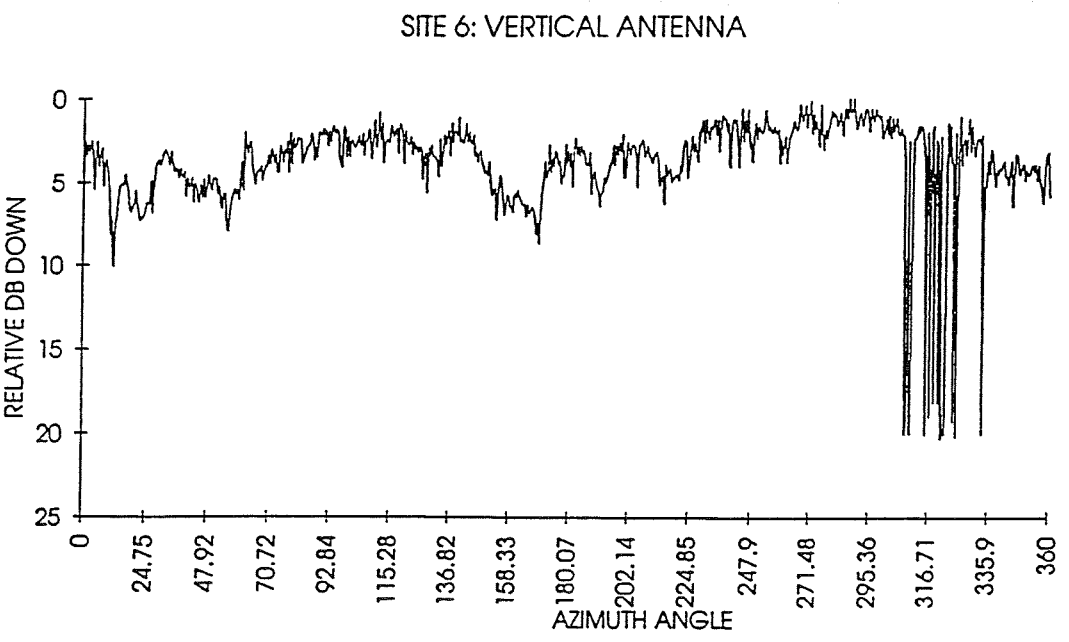
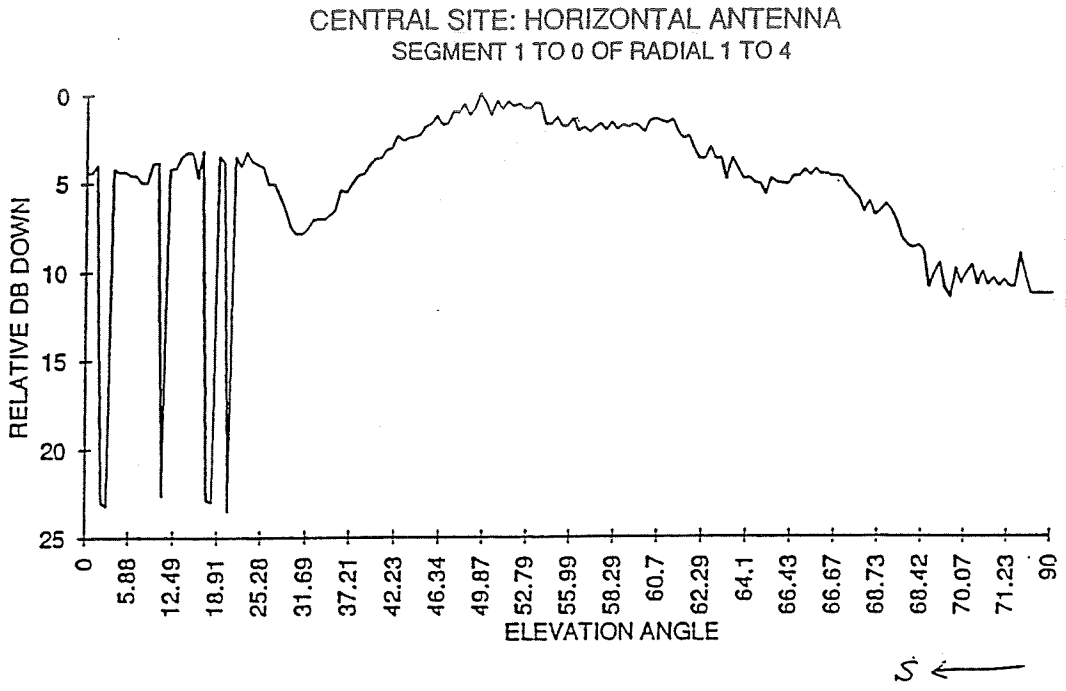


Figure 3-7: Site 6 azimuthal pattern. (30 degree elevation angle)

IV. ELEVATION PATTERNS OF THE ANTENNAS

The elevation patterns of the antennas were computed using the data obtained from the radial flights over the antenna sites. Referring to Figure 1-1, the flight paths were flown staring from site 1 to site 4, clover leaf from 4 to 3, radial from 3 to 6, a clover leaf from 6 to 5 and finally the radial from 5 to 2. All the radials passed over the central site. The elevation patterns for the central site obtained for the horizontal antenna are shown in Figures 4-1 to 4-3 and for the vertical antenna in Figures 4-4 to 4-6. For each radial path the results are presented in two segments. ----> S indicates the aircraft approaching the site and S ----> indicates the aircraft leaving the site. The elevation pattern for a circularly polarized signal for these antennas is expected to be symmetric, with a 3 db beamwidth of about a 144 degree and a variation in amplitude of about 12 db. In our case we had the aircraft flying over the antenna in one direction only, i.e. the aircraft did not cross the site again in two opposite directions. Instead we had three separate radial paths crossing the central site and this data was used to obtain complete information on the elevation pattern of the antenna. Elevation patterns for sites 1 and 4 were also computed using the radial from site 1 to site 4 for the two antennas. Figures 4-7 and 4-8 show the elevation pattern for site 1 for the horizontal and vertical antennas respectively and Figures 4-9 and 4-10 show the patterns for the site 4. In these cases also the sharp null in the pattern are due to problems with the transmitter. Sites 1 and 4 appear to have a more acceptable elevation angle variation as compared to the antenna at the central site.



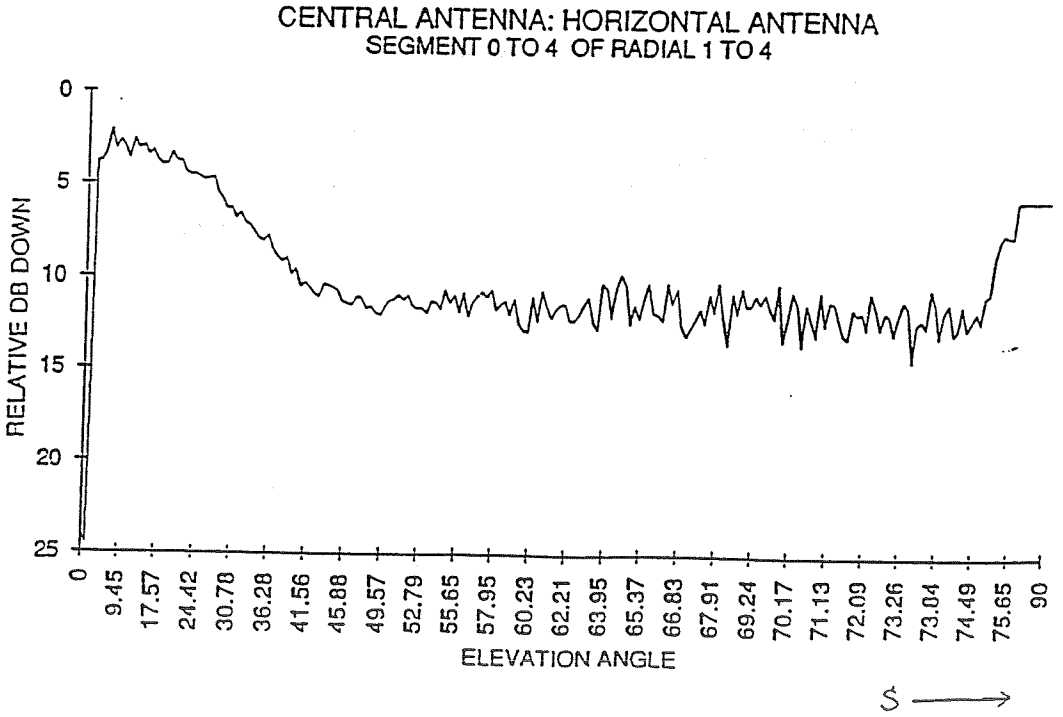
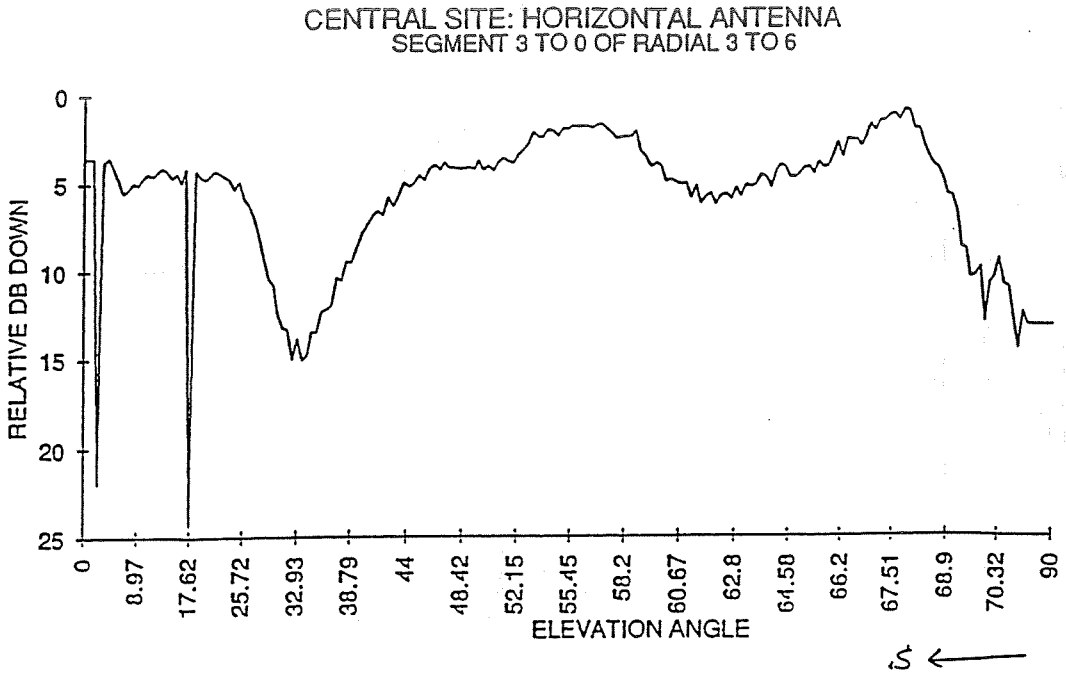


Figure 4-1: Central site elevation pattern; Radial from site 1 to 4 - Horizontal antenna.



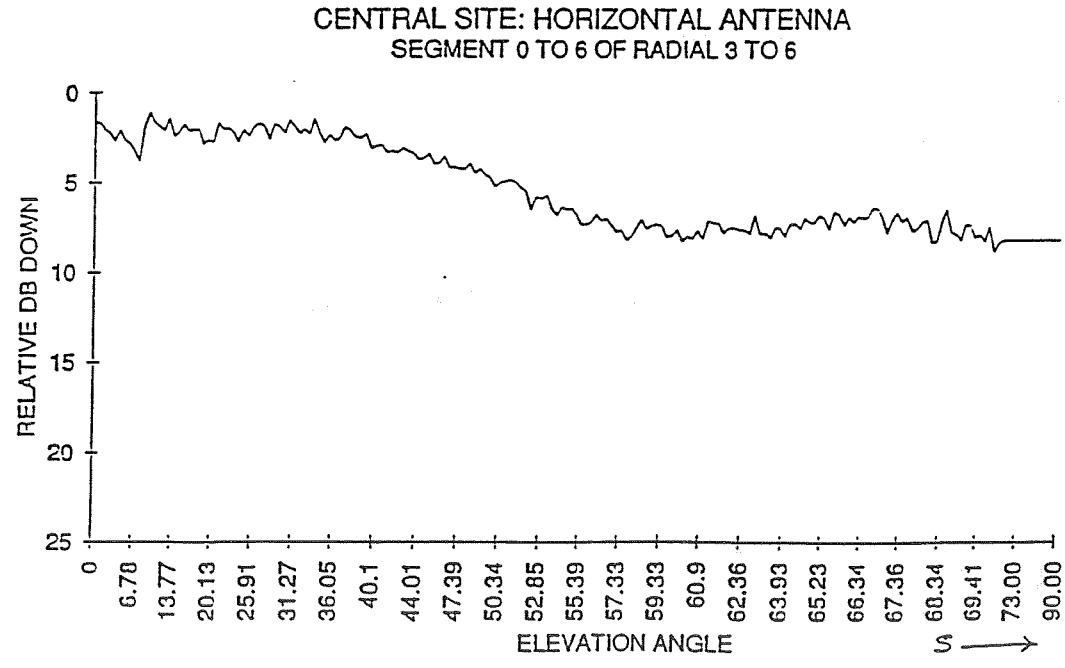
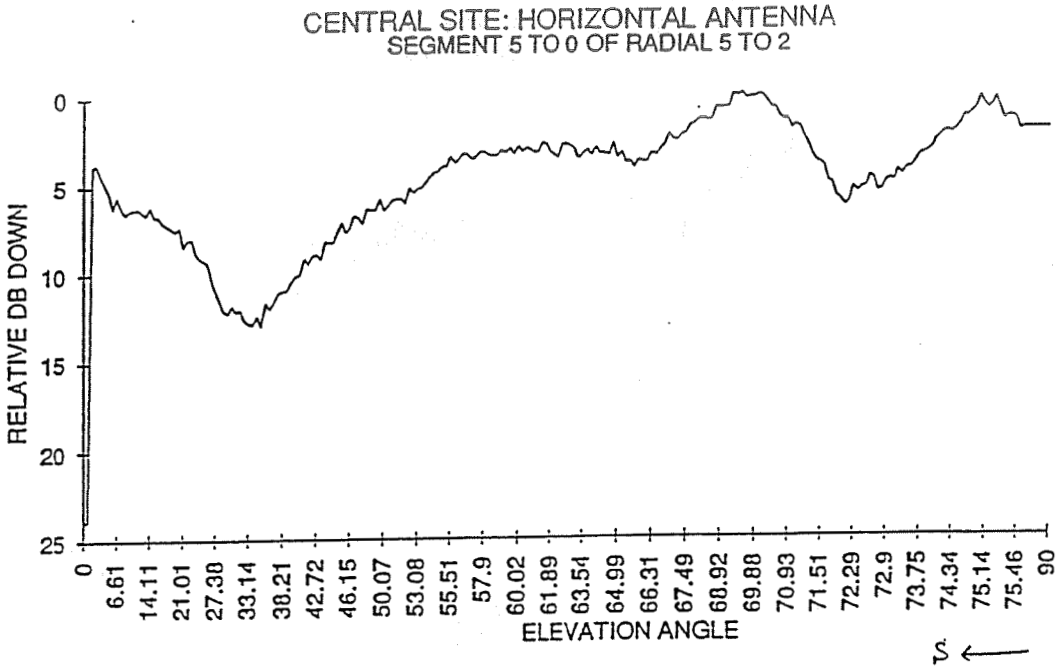


Figure 4-2: Central site elevation pattern; Radial from site 3 to 6 - Horizontal antenna.



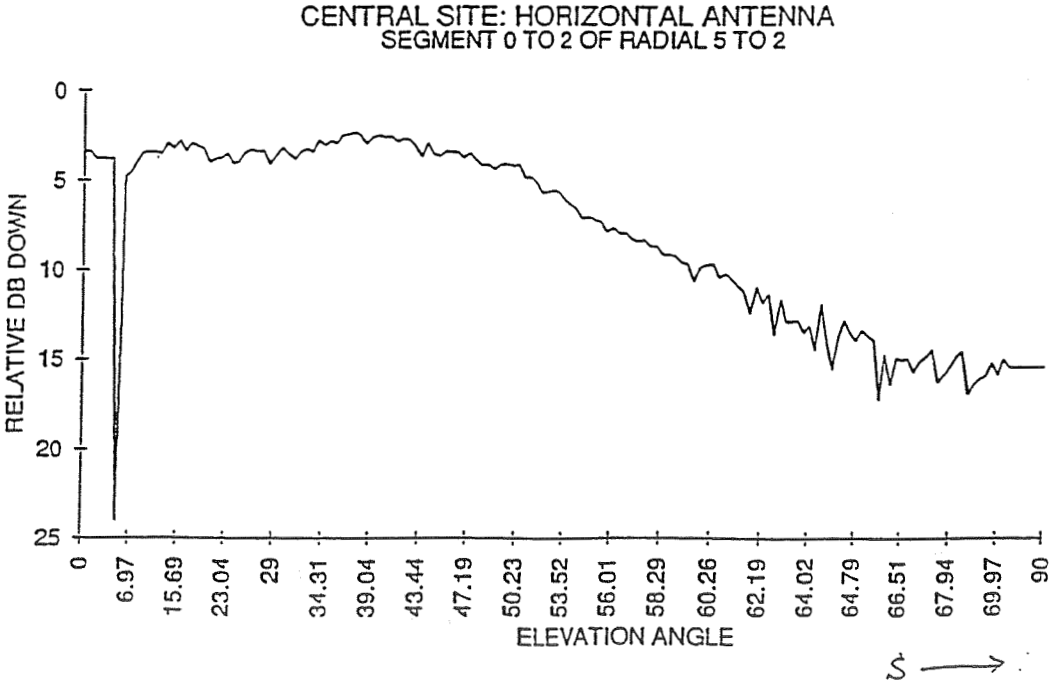
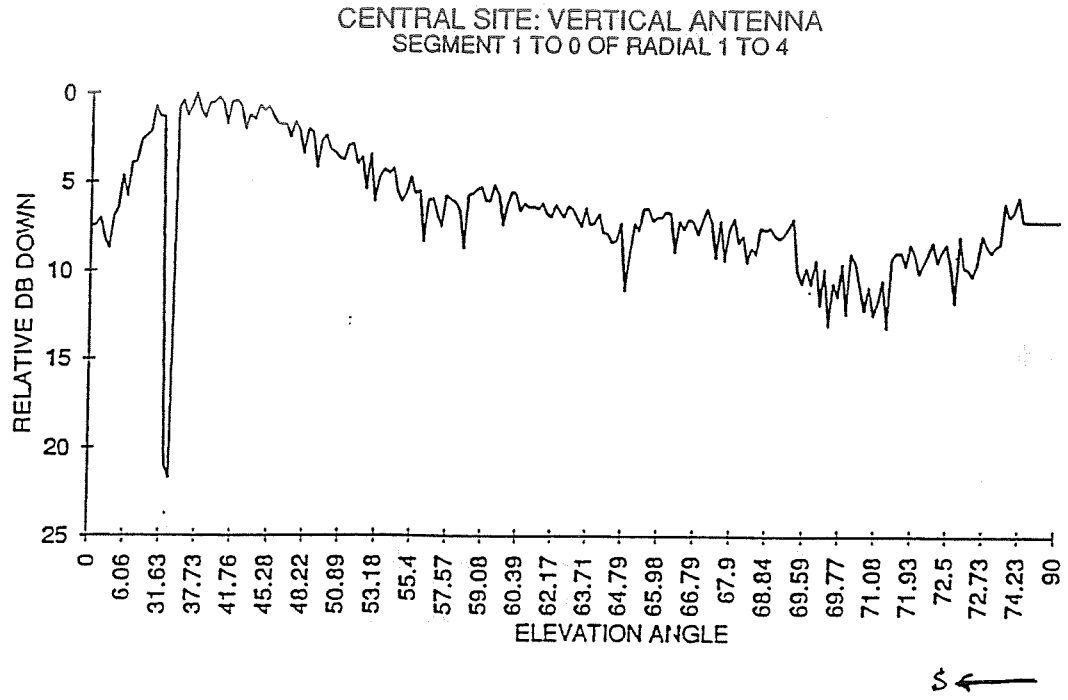


Figure 4-3: Central site elevation pattern; Radial from site 5 to 2 - Horizontal antenna.



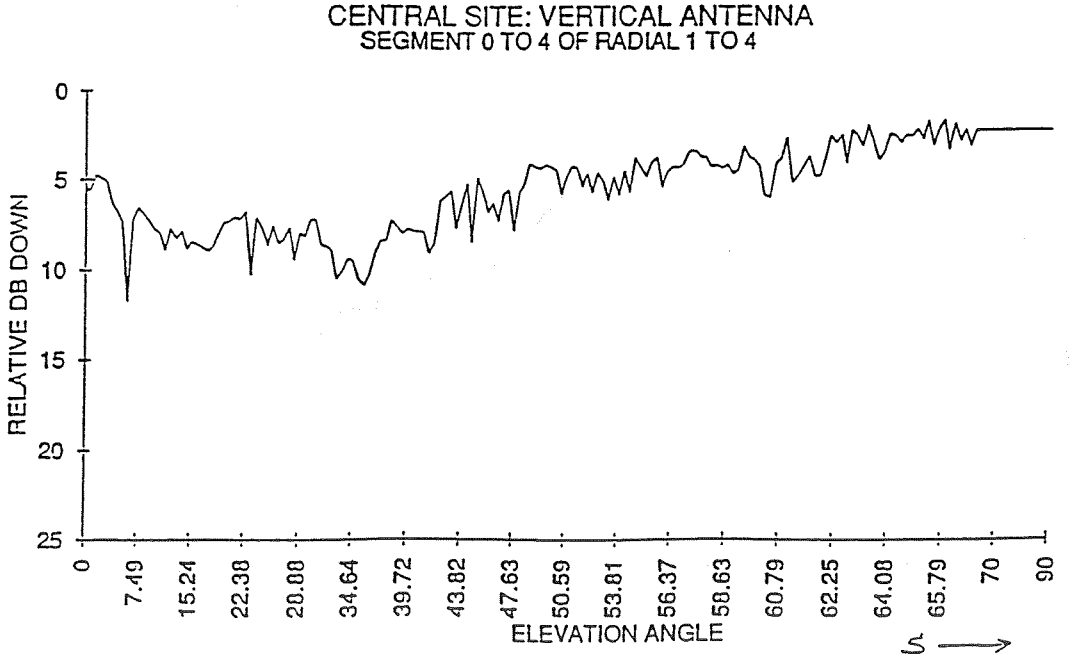
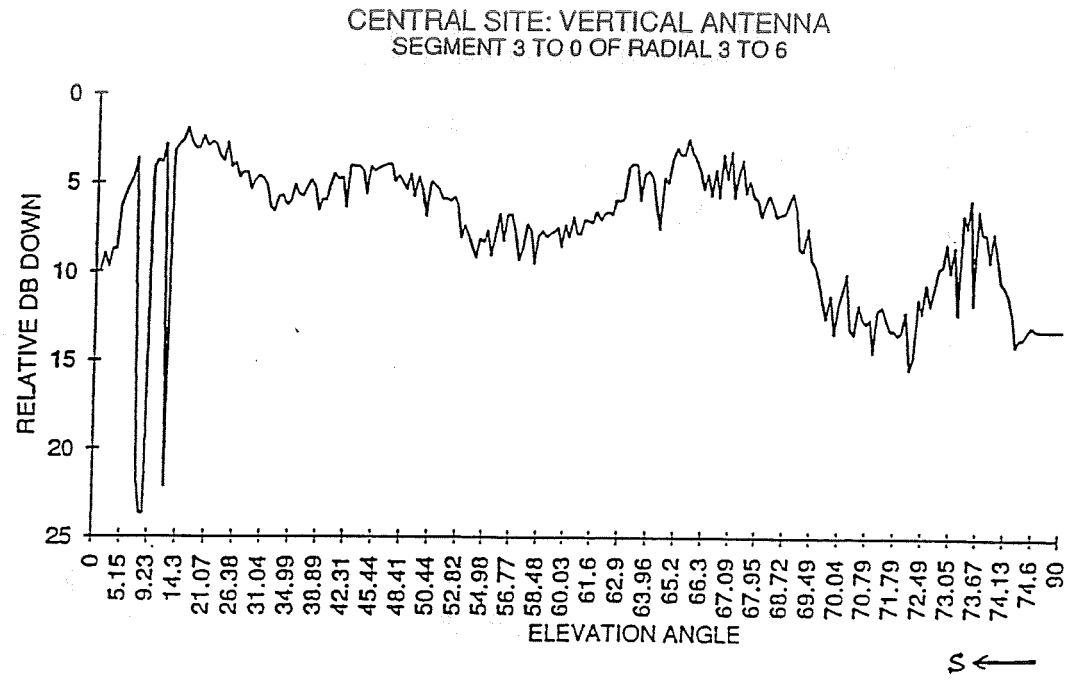


Figure 4-4: Central site elevation pattern; Radial from site 1 to 4 - Vertical antenna.



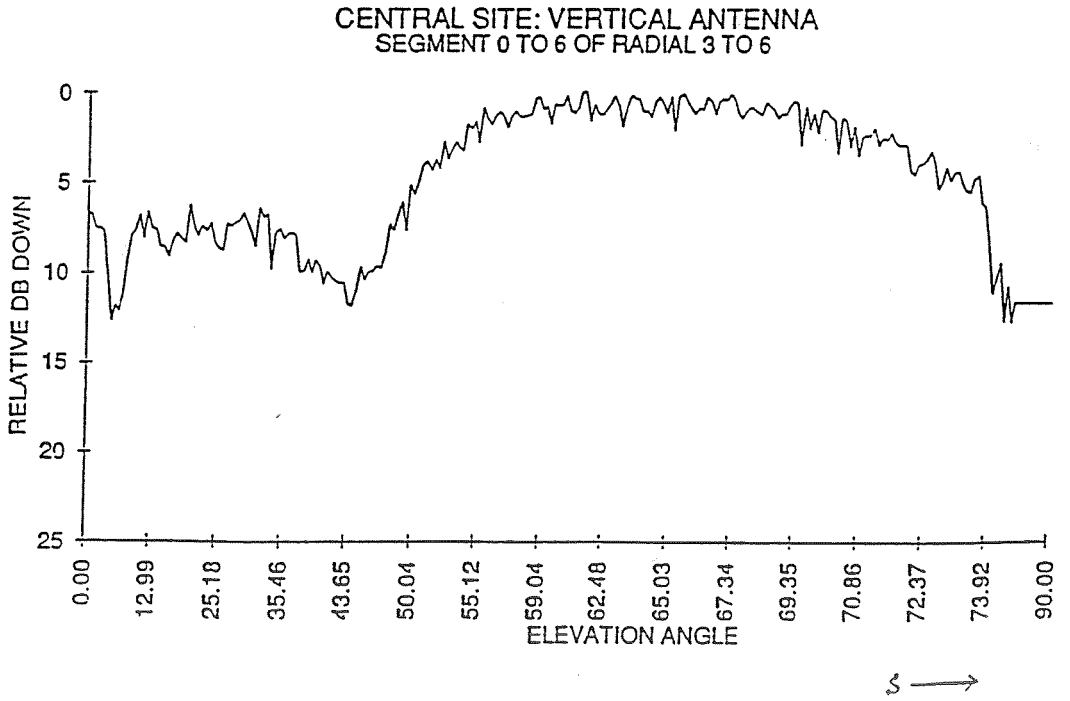
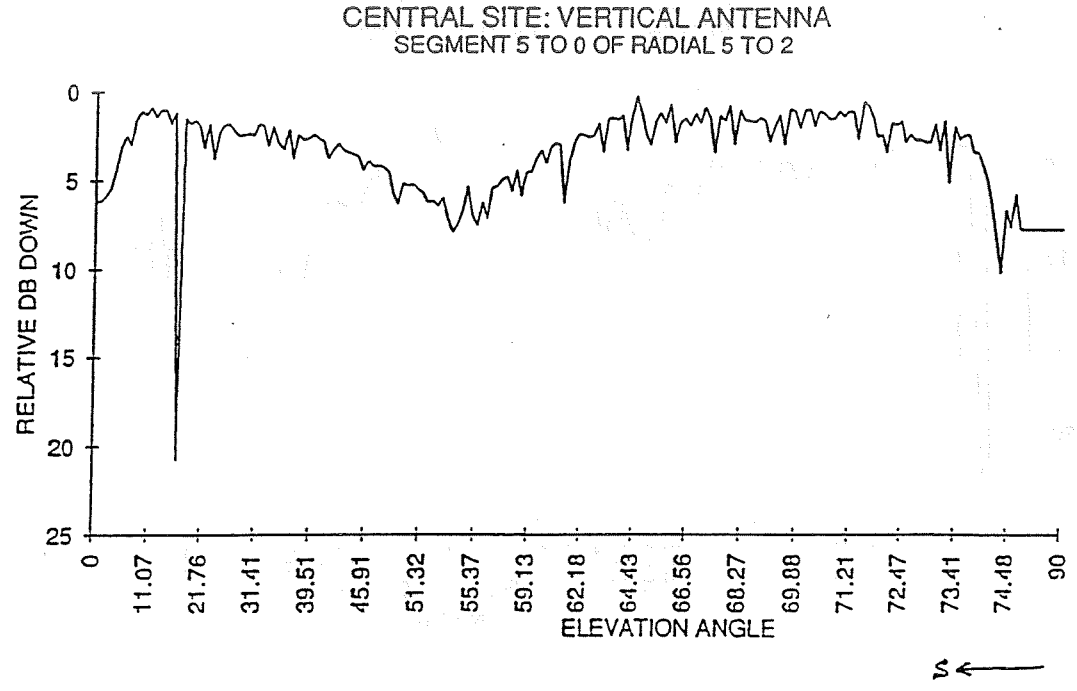


Figure 4-5: Central site elevation pattern; Radial from site 3 to 6 - Vertical antenna.



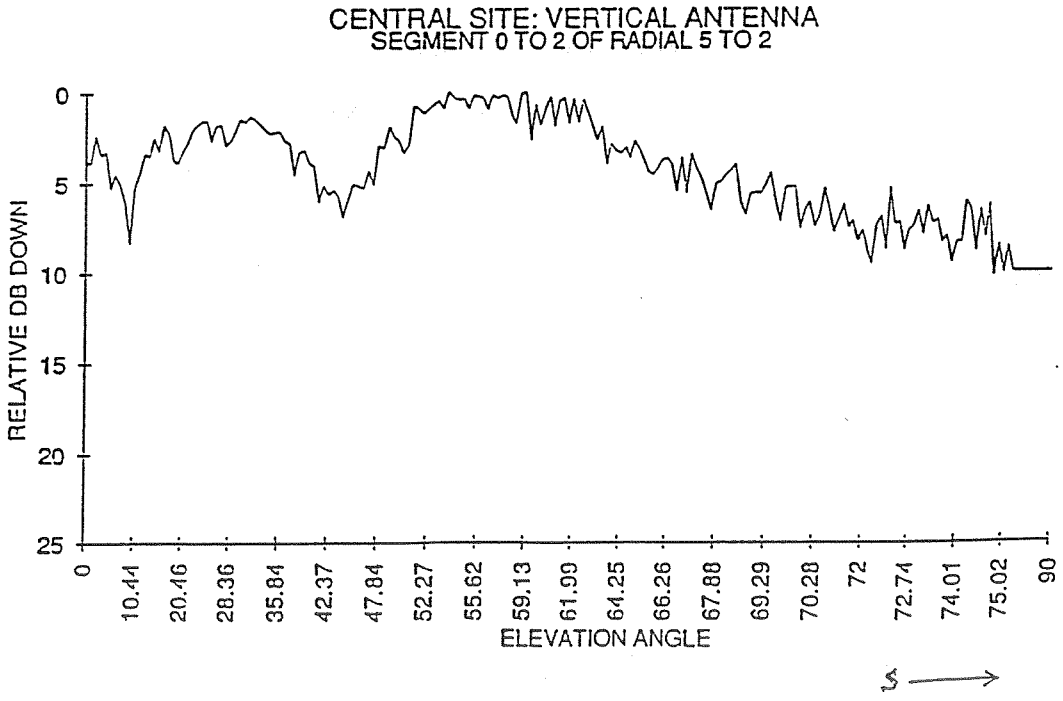
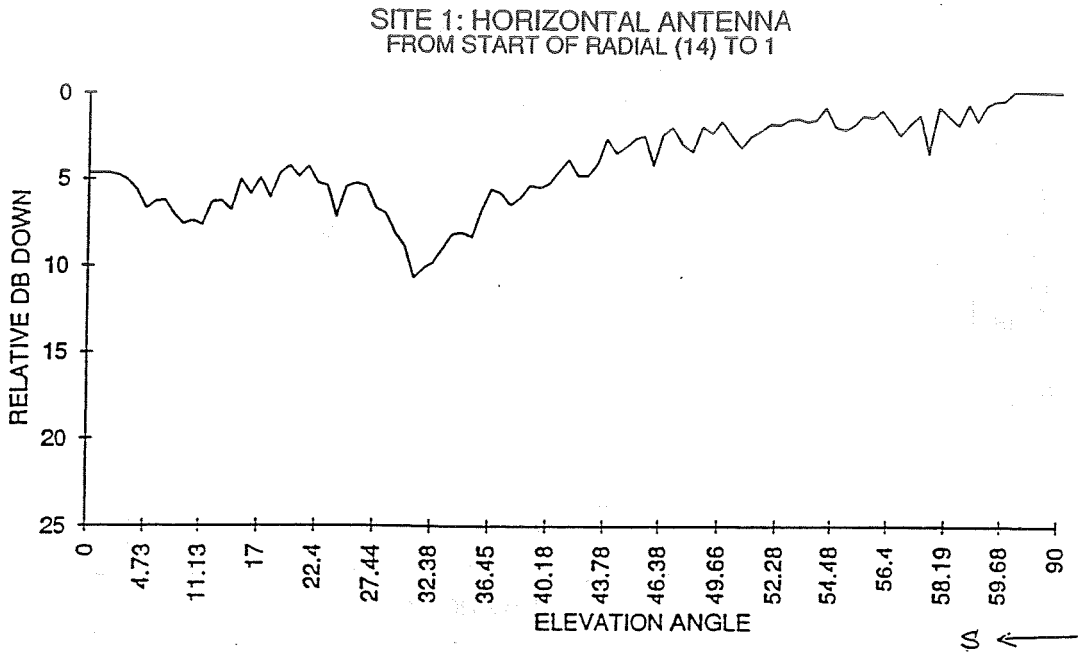


Figure 4-6: Central site elevation pattern; Radial from site 5 to 2 - Vertical antenna.



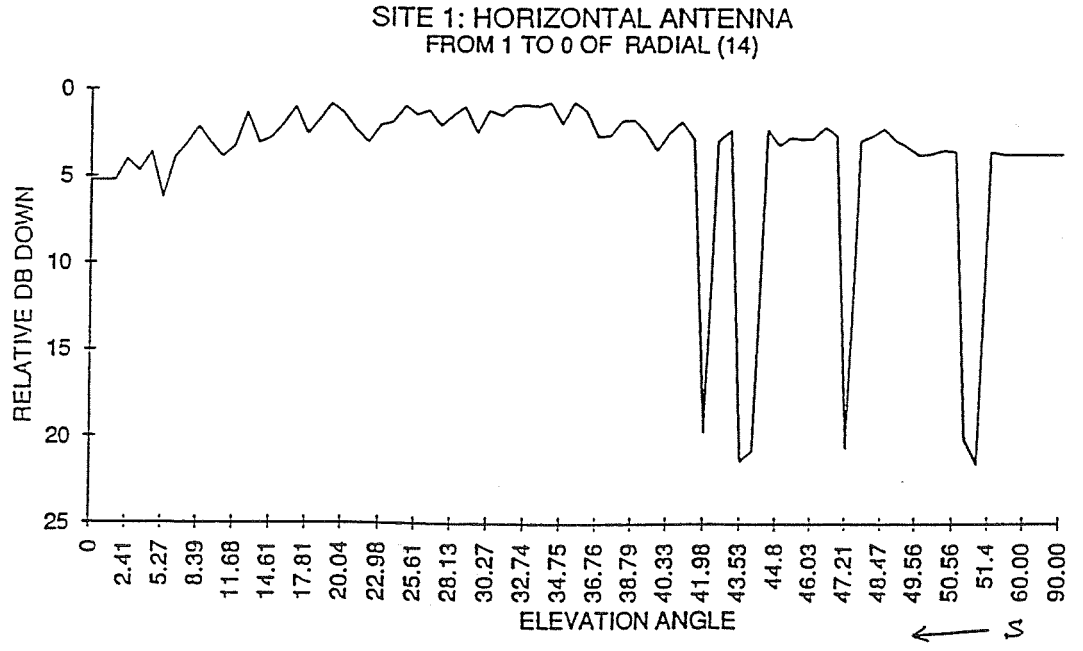


Figure 4-7: Site 1 elevation pattern; Radial from site 1 to 4 - Horizontal antenna.

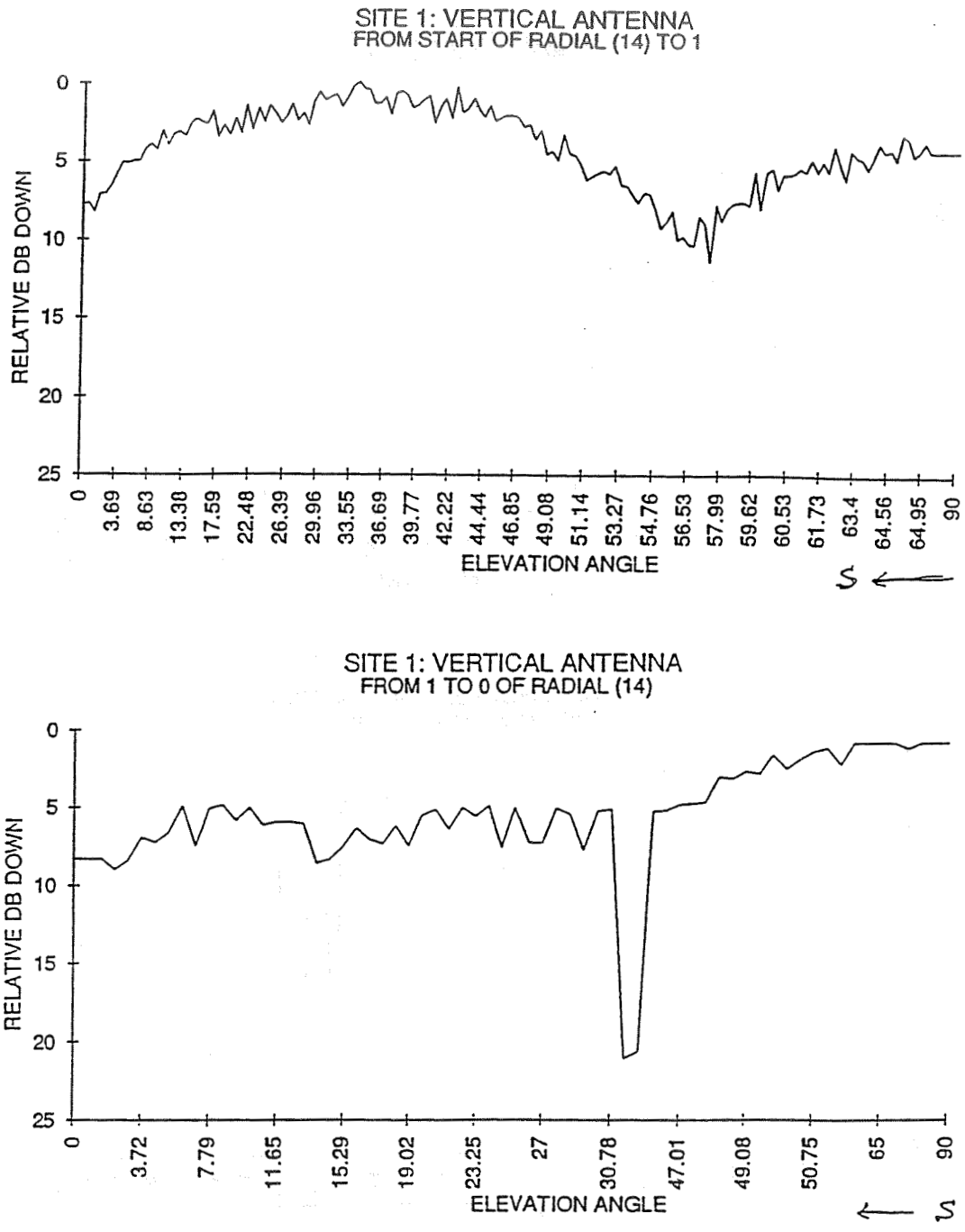
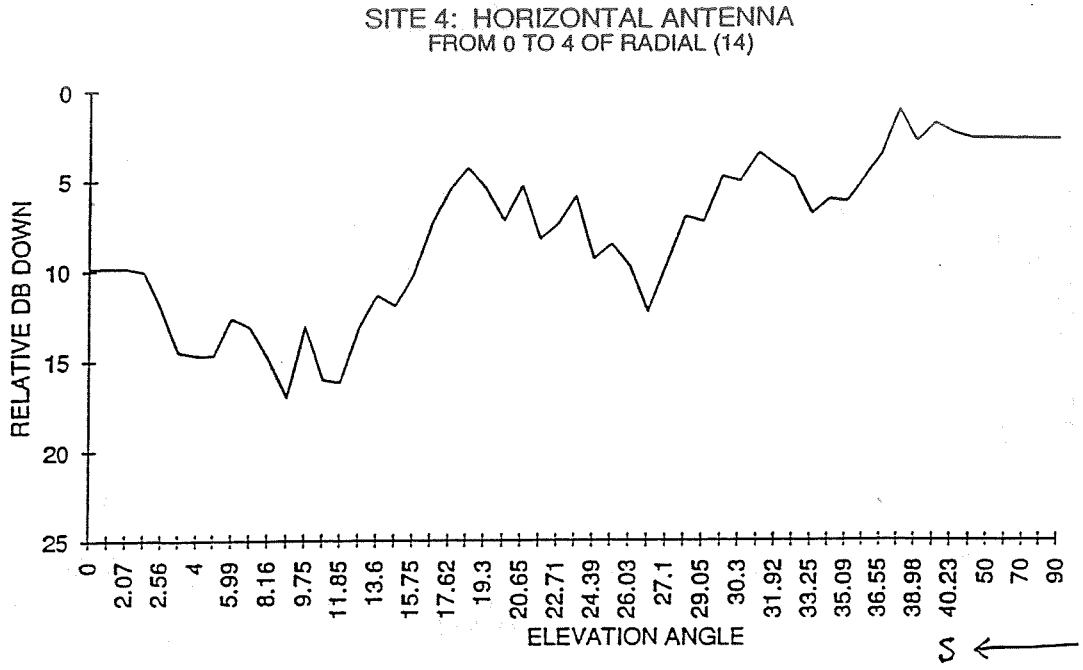


Figure 4-8: Site 1 elevation pattern; Radial from site 1 to 4 - Vertical antenna.



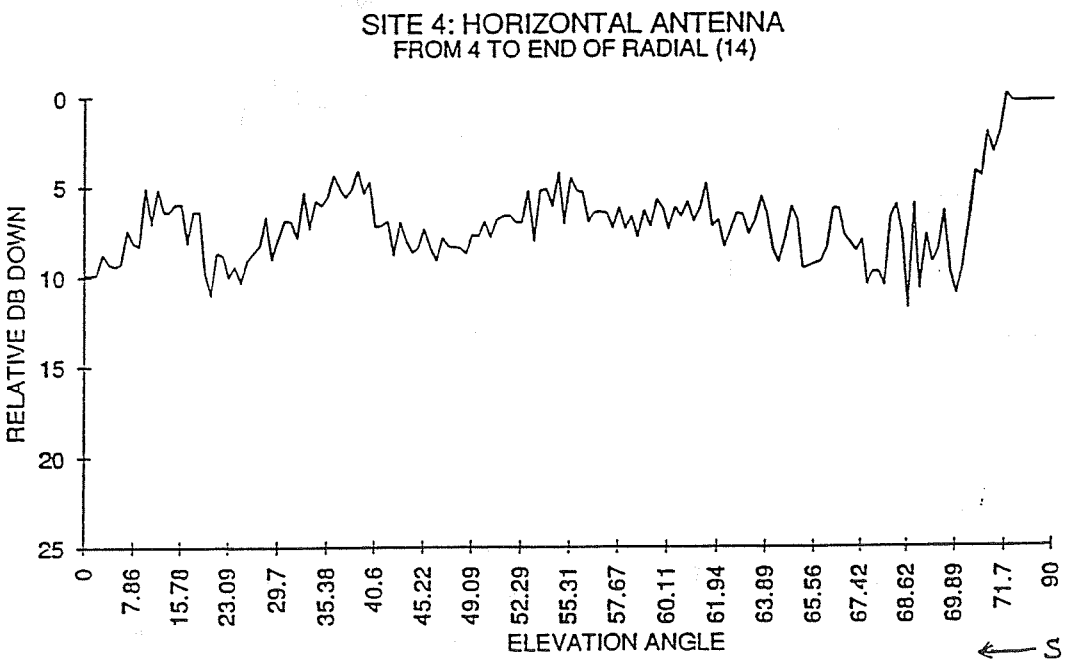


Figure 4-9: Site 4 elevation pattern; Radial from site 1 to 4 - Horizontal antenna.

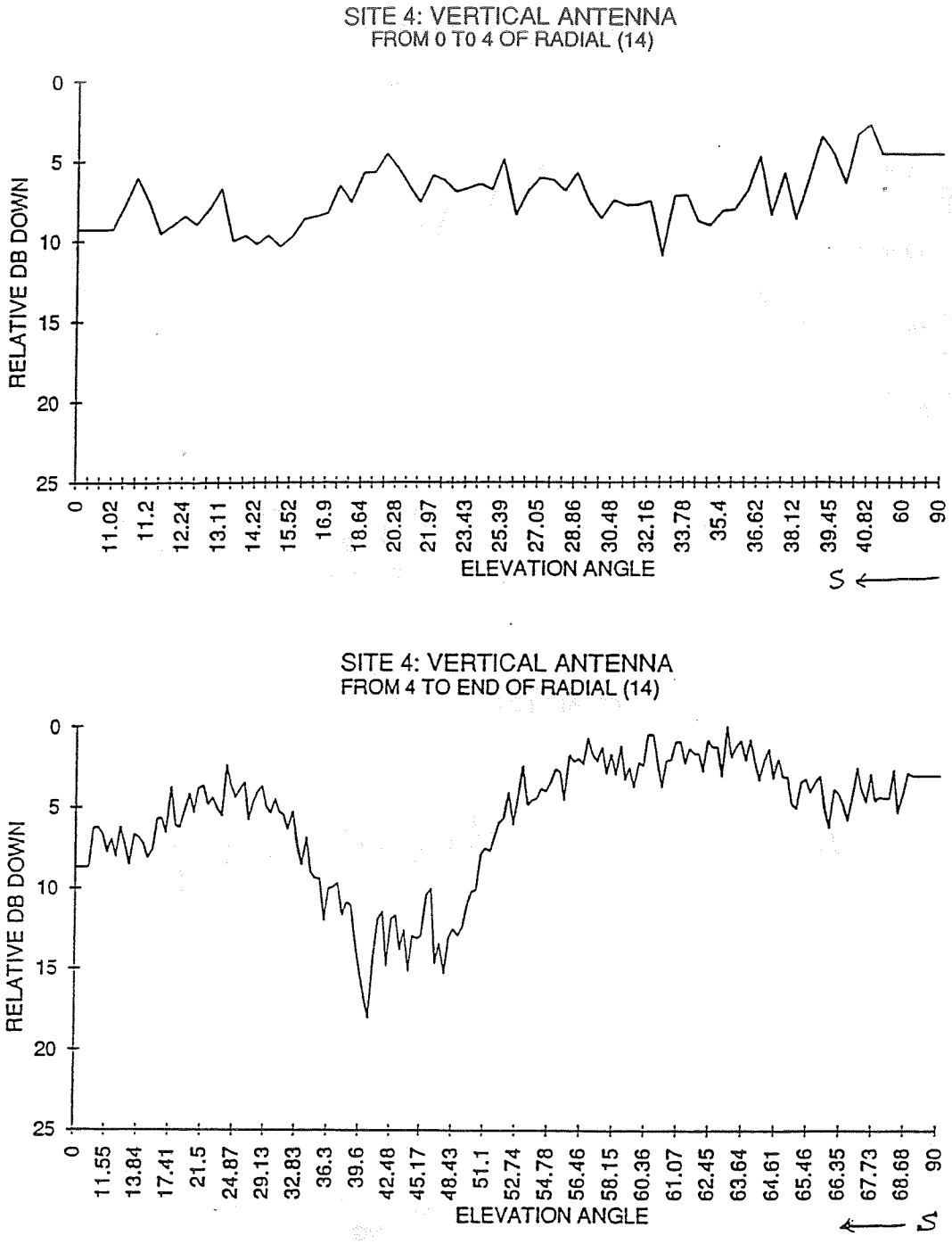


Figure 4-10: Site 4 elevation pattern; Radial from site 1 to 4 - Vertical antenna.

V. CONCLUDING REMARKS

Investigators have for many years been attempting to understand the complex phenomenon of the electromagnetics of lightning. Many models have been put forth, but due to the complex nature of the formation of the lightning channel many question are still left unanswered. The LDAR system at NASA, KSC offers a unique means and opportunity to track the lightning activity and to record the signal levels from the lightning sources. From a knowledge of the LDAR antenna patterns and using information on the signal levels of the lightning generated RF sources, it is possible to model the antenna pattern of the lightning source. The research reported here was undertaken with this goal in mind. The author was able to compute the antenna patterns for the LDAR system for vertical and horizontal polarization. The author proposes to continue this work with a graduate student, to analyze the lightning data that is already available. This continued work should lead to the determination of the effective radiated power of lightning sources and to the evaluation of a model for the antenna pattern of the lightning source.

REFERENCES

1. Maier, L. , C. Lennon and T. Britt: Lightning detection and Ranging system performance analysis. Proceedings, Sixth Conference on Aviation Weather Systems, 75th American Meteorological Society's Meeting, Dallas, Texas, Jan. 1995.

2. Lennon, C. , L. Maier: Lightning mapping system, NASA CP-3106, Vol. II, Proceedings, 1991 International Aerospace and Ground Conference on Lightning and Static Electricity, 1991

1995 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

S20-18
7760
p. 18

EFFECT OF ENTRY OF SUBCOOLED CRYOGEN ON THERMAL STRATIFICATION IN A CRYOGENIC STORAGE TANK

Mr. Pao-lien Wang
Associate Professor
Department of Engineering Technology
University of North Carolina at Charlotte
Charlotte, North Carolina

KSC Colleague - Frank Howard
Fluids

Contract Number NASA-NGT-60002
Supplement 19

July 28, 1995

ACKNOWLEDGMENTS

I would like to thank NASA/ASEE Summer Faculty Fellowship Program, Mr. William C. Jones of Kennedy Space Center, and Dr. E. Ramon Hosler of the University of Central Florida, for offering me with the opportunity to work on this challenging and rewarding project. Special thanks should be extended to my colleague Mr. Frank S. Howard for his help, guidance, and encouragement throughout the project. The friendliness and helpfulness of Ms. Kari Stiles is also very much appreciated.

ABSTRACT

The purpose of this study was to predict if subcooled cryogenic liquid entering the bottom of a storage tank will destroy the thermal stratification of the tank.

After an extensive literature search, a formula for maximum critical Reynolds Number which used to predict the destratification of a cryogenic tank was found. Example of calculations and graphics to determine the mixing of fluid in the tank were presented.

SUMMARY

1. Formula of Critical Reynolds number to predict destratification was shown.

2. Graphs of Reynolds number of inlet flow pipe, and critical Reynolds number versus different inlet pipe diameter were presented.

3. A sample of calculation of liquid hydrogen tank, and conclusion was given.

TABLE OF CONTENTS

1. INTRODUCTION

1.1 BACKGROUND

Thermal stratification exists in cryogenic fluid in a storage tank when there is a temperature differential between fluids on the top and the bottom of the storage tank. This temperature differential is caused by the buoyant forces resulting from the density difference between hot and cold fluid. Because of the buoyant forces, it is evident that some degree of stratification exist in almost all cryogenic storage tanks unless there is a mixing force.

This study involved a cryogenic refrigeration subcooling systems which take liquid from the top of a cryogenic storage tank and passed it through a refrigerator and back into the bottom of the tank. Figure 1 shows a schematic of piping connecting the storage tank. In order to make the refrigeration process more efficient the system should provide as high a temperature difference between the cryogen at the inlet and outlet of the refrigerator as possible. Mixing of the fluid in the tank would lower the temperature of the cryogen going into the refrigerator and lower the temperature difference. This study was made to determine how to analyze the effect of subcooled liquid entering the bottom of the tank on thermal stratification in the tank.

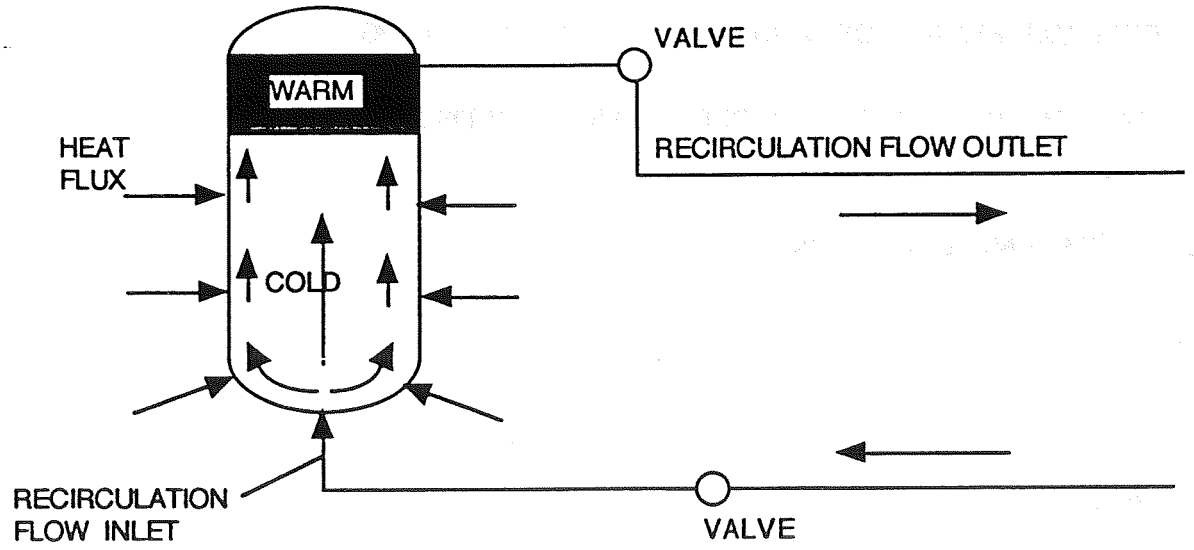


Fig. 1 Schematic of Cryogenic Tank System

1.2 OBJECTIVE

The object of this study was to gather applicable engineering information and data and see if cryogen entering the bottom of a cryogenic storage tank, will cause the hot and cold liquid to mix. The cryogenic liquids studied were liquid oxygen and liquid hydrogen. Typical tank sizes were selected.

2. CALCULATIONS

2.1 CRITICAL REYNOLDS NUMBER AND REYNOLDS NUMBER

A critical Reynolds number of the inlet flow pipe was defined in reference [3] as that value where the buoyant force of tank liquid becomes unimportant. For a Reynolds number greater than this critical value, the system may be assumed to mix completely. Below this critical Reynolds number, the buoyant forces may be strong enough to limit the degree of mixing. The critical Reynolds number was given by a correlation available in reference [3].

The critical Reynolds number is given by

$$Re^2 = 0.912\left(\frac{D}{H}\right)^{\frac{2}{3}} Pr^{\frac{-2}{3}} Gr^{*\frac{2}{3}} \qquad (1)$$

where Re = Critical Reynolds number of inlet pipe

Pr = Prandtl number

D = Tank diameter

H = Tank height

Gr* = Modified Grashof number

The modified Grashof number is given in reference [3] as

$$Gr^* = \frac{g\beta q'' L^4}{k\nu^2} \qquad (2)$$

where: g = Gravitational acceleration

β = Coefficient of thermal expansion of liquid

q" = Heat flux

L = Liquid height

k = Thermal conductivity of liquid

ν = Kinematic viscosity of liquid

The Reynolds number is given as

$$Re = \frac{Vd}{\nu^2} \qquad (3)$$

where V = Flow velocity

d = Inside diameter of inlet pipe

ν = Kinematic viscosity

2.2 SAMPLE OF CALCULATION

2.2.1 THE CRITICAL REYNOLDS NUMBER FOR LIQUID HYDROGEN

The critical Reynolds number for liquid hydrogen is calculated from data given in Table 1. Pressure, temperature, and heat flux are assumed.

TABLE 1 **LH2 Tank**

Term	Symbol	Unit	Value	Reference
Pressure in tank	p	psia	35	
Outlet temperature	To	R	27	
Tank diameter	D	ft.	10	
Tank height	H	ft.	37.75	
Tank volume	Vt	ft^3	2669	

TABLE 1 (Continued)

Term	Symbol	Unit	Value	Reference
Inside dia. of inlet pipe.	d	ft.	0.5	
Inlet flow rate	Q	ft^3 /hr.	1299.5	
Thermal expansion	β	1/R	0.006256	[1]
Density	ρ	lbm/ft^3	4.758	[1]
Specific heat	Cp	Btu/lbm. R	1.6595	[1]
Dynamic viscosity	μ	lbm/ft. hr.	0.0539	[1]
Thermal conductivity	k	Btu/ft.hr.R	0.04795	[1]
Prandtl number	Pr		1.867	[1]
Heat flux	q"	$Btu/hr.\ ft^2$	14.23	
Kinematic viscosity	ν	$ft.^2/hr.$	0.01133	[1]
Tank liquid height	L	ft.	34.5	
Tank liquid volume	Vl	ft^3	2562	
Gravitation acceleration	g	$ft./hr^2$	417,312,000	

From Equation (2)

$$Gr^* = \frac{g\beta L^4 q''}{k\nu^2} = \frac{4.173(10^8)\frac{ft}{hr^2}(6.256)(10)^{-3}\left(\frac{1}{R}\right)(34.5)^4\, ft^4(14.23)\frac{Btu}{ft^2 hr}}{4.795(10)^{-2}\frac{Btu}{hr.ft.R}\left((1.133)(10)^{-2}\right)^2\frac{ft^4}{hr^2}}$$

$Gr^* = 8.549 \times 10^{18}$

From Equation (1)

$$Re^2 = 0.912\left(\frac{D}{H}\right)^{\frac{2}{3}} Pr^{\frac{-2}{3}} Gr^{*\frac{2}{3}}$$

$$Re = \left[0.912\left(\frac{10}{37.75}\right)^{\frac{2}{3}}(1.867)^{\frac{-2}{3}}\left[8.549(10)^{18}\right]^{\frac{2}{3}}\right]^{\frac{1}{2}}$$

Re (Critical Reynolds number) = 1.0256×10^6

(for the selected tank with D = 10 ft, H = 37.75 ft and L = 34.5 ft.)

2.2.2 THE CRITICAL REYNOLDS NUMBER AND REYNOLDS NUMBER FOR LIQUID OXYGEN

The calculation for Critical Reynolds number and Reynolds numbers at different inlet pipe diameters are shown in the Appendix section.

3. RESULTS AND CONCLUSION

3.1 RESULTS

The Critical Reynolds number and Reynolds number for both liquid hydrogen and liquid oxygen are calculated by using the Microsoft Excel spreadsheet are completed (see Appendix). Graphs of critical Reynolds number and Reynolds number versus inlet pipe diameter for both LH2 and LO2 are shown in Fig. 2 and Fig. 3.

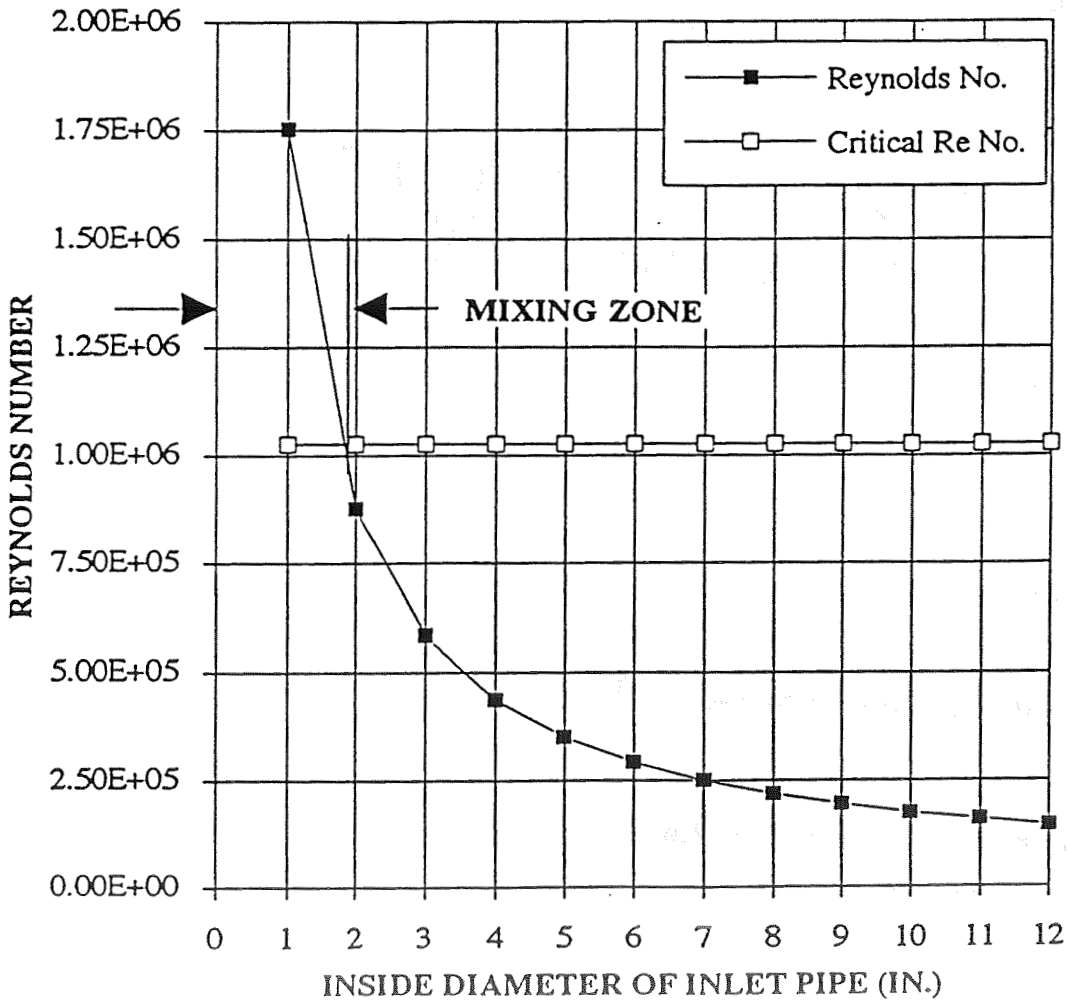


Fig. 2 Liquid Hydrogen Tank Mixing Curve

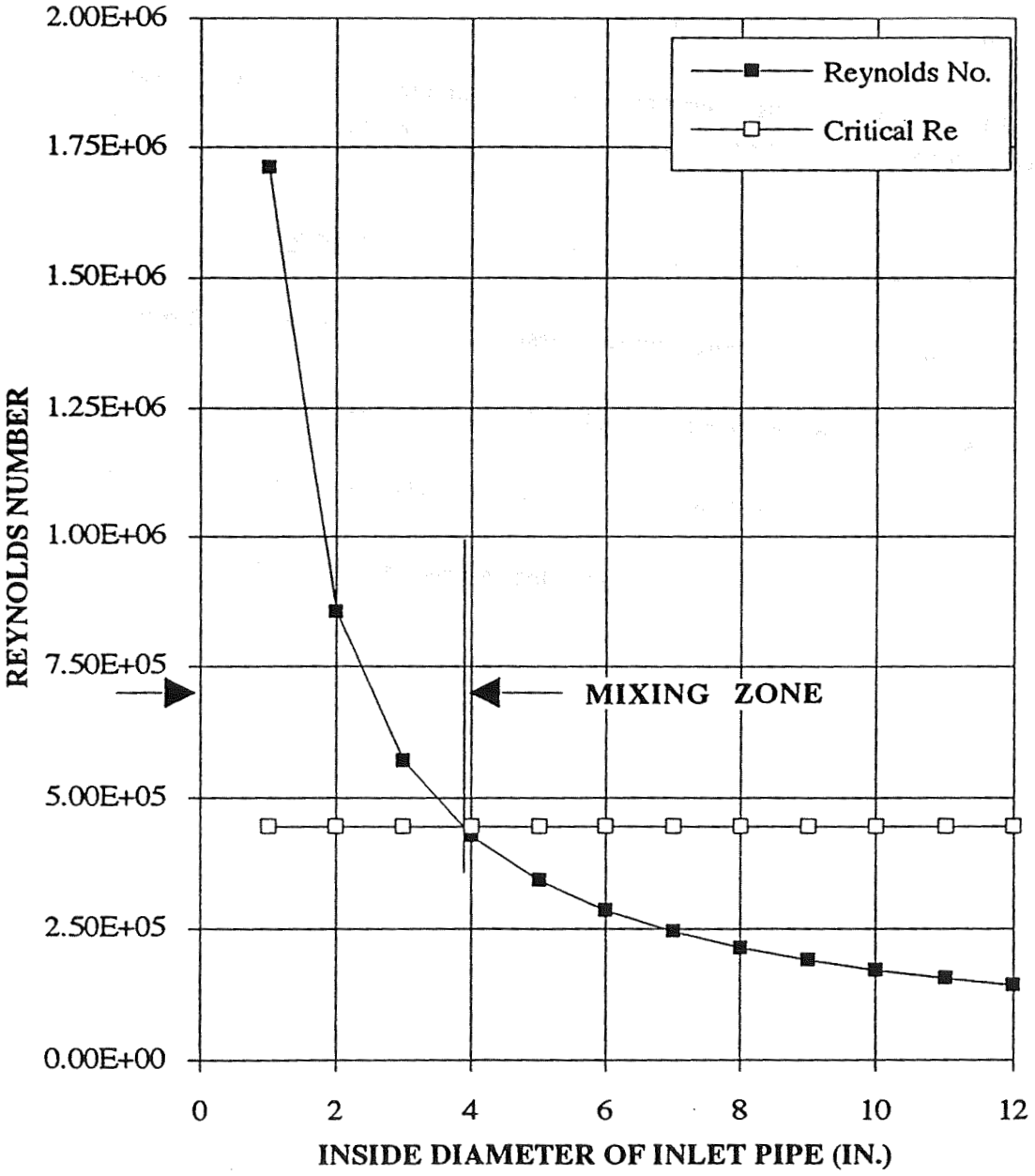


Fig. 3 Liquid Oxygen Tank Mixing Curve

3.2 CONCLUSION

In conclusion, form Fig. 2, it indicates that for liquid hydrogen tank an inlet pipe diameter of about 1.8 in. or less, the tank fluid will have a complete mix. Diameter of the inlet pipe greater the 1.8 in will be large enough to prevent complete mixing.

Fig. 3 indicates that for liquid oxygen tank an inlet pipe diameter of about 4 in. or less, the tank fluid will have a complete mix. Diameter of the inlet pipe greater the 4 in. will prevent complete mixing.

The dimensions of the selected hydrogen tank are D = 10 ft., H = 37.75 ft., and L = 34.5 ft.

The pressure and temperature at the tank inlet were assumed to be 25 psia and 27 R respectively.

The dimensions of the selected oxygen tank are D = 10 ft., H = 23.83 ft., and L = 21.22 ft.

The pressure and temperature at the tank inlet were assumed to be 35 psia and 141 R respectively.

Results shown in this report are used as a guide to place a reasonable prediction only.

REFERENCES

[1] Robert D. McCarty. *Hydrogen Technological Survey - Thermophysical properties.* NASA S-3089, National Aeronautics and Space Administration, Washington, D.C. 1975.

[2] Hans M. Roder and Lloyd A. Weber. *ASRDI Oxygen Technology Survey, Volume I: Thermophysical Properties,* NASA SP-3071, National Aeronautics and Space Administration, Washington, D.C. 1972.

[3] T. M. Lovrich and S. H. Schwartz. *Development of Thermal Stratification and Destratification Scaling Concepts - Volume I, Definition of Thermal Stratification Scaling Parameters and Experimental Investigations*, NASA CR-143944, NASA George C. Marshall Space Center, Marshall Space Center, Alabama, October 1975.

APPENDIX

CALCULATION OF MODIFIED GRASHOF NUMBER

Formula:

$$GR^* = \frac{g\beta q'' L^4}{k\nu^2}$$

CALCULATION OF CRITICAL REYNOLDS NUMBER

Formula:

$$\mathrm{Re}^2 = 0.912\left(\frac{D}{H}\right)^{\frac{2}{3}} \mathrm{Pr}^{\frac{-2}{3}} Gr^{*\frac{2}{3}}$$

TABLE 1 LIQUID HYDROGEN TANK

Term	Symbol	Unit	Value	Reference
Pressure	p	psia	35	
Outlet temperature	To	R	27	
Tank diameter	D	ft.	10	
Tank height	H	ft.	37.75	
Tank volume	V	cu ft.	2669	
Inside dia. of inlet pipe.	d	ft.	0.5	
Inlet flow rate	Q	cu ft./hr.	1299.5	
Thermal expansion coefficient	β	1/R	0.006256	[1]
Density	ρ	lbm/cu ft.	4.758	[1]
Specific heat	Cp	Btu/lbm. R	1.6595	[1]
Dynamic viscosity	μ	lbm/ft. hr.	0.0539	[1]
Thermal conductivity	k	Btu/ft.hr.R	0.04795	[1]
Prandtl number	Pr		1.867	[1]
Heat flux	q"	Btu/hr. sq ft.	14.23	
Kinematic viscosity	ν	sq ft./hr.	0.01133	[1]
Tank liquid height	L	ft.	34.5	
Tank liquid volume	Vt	cu ft	2562	
Gravitation acceleration	g	ft./sq hr	417,312,000	

LH2 Tank

Modified Grashof number

$$Gr^* = 8.549 \times 10^{18}$$

Critical Reynolds number of the LH2 tank with selected D, H, and L

$$Re = 1.0256 \times 10^{6}$$

Data for curve in Fig. 2

Inside Pipe Diameter (in.)	Area (sq ft.)	velocity (ft/s)	Reynolds No.	Critical Re No.
1	5.45E-03	6.62E+01	1.75E+06	1.03E+06
2	2.18E-02	1.65E+01	8.76E+05	1.03E+06
3	4.91E-02	7.35E+00	5.84E+05	1.03E+06
4	8.73E-02	4.14E+00	4.38E+05	1.03E+06
5	1.36E-01	2.65E+00	3.50E+05	1.03E+06
6	1.96E-01	1.84E+00	2.92E+05	1.03E+06
7	2.67E-01	1.35E+00	2.50E+05	1.03E+06
8	3.49E-01	1.03E+00	2.19E+05	1.03E+06
9	4.42E-01	8.17E-01	1.95E+05	1.03E+06
10	5.45E-01	6.62E-01	1.75E+05	1.03E+06
11	6.60E-01	5.47E-01	1.59E+05	1.03E+06
12	7.85E-01	4.60E-01	1.46E+05	1.03E+06

TABLE 2 LIQUID OXYGEN TANK

Term	Symbol	Unit	Value	Reference
Pressure	p	psia	35	
Outlet temperature	To	R	140.5	
Tank diameter	D	ft.	10	
Tank height	H	ft.	23.83	
Tank volume	V	cu ft.	1589	
Inside dia. of inlet pipe	d	ft.	0.4	
Tank inlet flow rate	Q	Cu ft./hr.	1026.74	
Thermal expansion coefficient	β	1/R	0.02167	[2]
Density	ρ	lbm/cu ft.	74.99	[2]
Specific heat	Cp	Btu/lbm. R.	0.4	[2]
Dynamic viscosity	μ	lbm./ft. hr.	0.687	[2]
Thermal conductivity	k	Btu/ft.hr.R.	0.09685	[2]
Prandtl number	Pr		2.837	[2]
Heat flux	q"	Btu/hr.sq ft.	29.681	
Kinematic viscosity	ν	Sq ft./hr.	0.00916	[2]
Tank liquid height	L	ft.	21.215	
Tank liquid volume	Vl	Cu ft.	1525.44	
Gravitation acceleration	g	ft./sq hr	417,312,000	

LO2 Tank

Modified Grashof number

$$Gr^* = 6.691 \times 10^{17}$$

Critical Reynolds number of LO2 tank with selected D, H, and L

$$Re = 4.4468 \times 10^{5}$$

Data for curve in Fig. 3

Inside Pipe Diameter (in.)	Area (sq ft.)	Velocity ft./s	Reynolds No.	Critical Re
1	5.45E-03	5.23E+01	1.71E+06	4.45E+05
2	2.18E-02	1.31E+01	8.56E+05	4.45E+05
3	4.91E-02	5.81E+00	5.71E+05	4.45E+05
4	8.73E-02	3.27E+00	4.28E+05	4.45E+05
5	1.36E-01	2.09E+00	3.43E+05	4.45E+05
6	1.96E-01	1.45E+00	2.85E+05	4.45E+05
7	2.67E-01	1.07E+00	2.45E+05	4.45E+05
8	3.49E-01	8.17E-01	2.14E+05	4.45E+05
9	4.42E-01	6.46E-01	1.90E+05	4.45E+05
10	5.45E-01	5.23E-01	1.71E+05	4.45E+05
11	6.60E-01	4.32E-01	1.56E+05	4.45E+05
12	7.85E-01	3.63E-01	1.43E+05	4.45E+05

1995 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

S21-61
7761
P. 26

THE USE OF A BLOCK DIAGRAM SIMULATION LANGUAGE FOR RAPID MODEL PROTOTYPING

Dr. Jonathan E. Whitlow
Associate Professor
Chemical Engineering Department
Florida Institute of Technology
Melbourne, Florida

KSC Colleague - Peter Engrand
Computer Science

Contract Number NASA-NGT-60002
Supplement 19

August 18, 1995

ACKNOWLEDGEMENTS

I would like to express gratitude to my NASA colleague Peter Engrand for the opportunity to participate in the Summer Faculty Research Program. In addition, thanks are owed to the Advanced Software group, in particular I am indebted to Charlie Goodrich and Bob Merchant for their help and technical support during my stay. I would also like to extend compliments to Dr. Ray Hosler and Kari Stiles of UCF as well as Greg Buckingham of NASA for providing a professional and stimulating summer program.

ABSTRACT

The research performed this summer focused on the development of a predictive model for the loading of liquid oxygen (LO_2) into the external tank (ET) of the shuttle prior to launch. A predictive models can greatly aid the operational personnel since instrumentation aboard the orbiter and ET is limited due to weight constraints. The model, which focuses primarily on the orbiter section of the system was developed was developed using a block diagram based simulation language known as VisSim. Simulations were run on LO_2 loading data for shuttle flights STS50 and STS55 and the model was demonstrated to accurately predict the sensor data recorded for these flights. As a consequence of the simulation results, it can be concluded that the software tool can be very useful for rapid prototyping of complex models.

TABLE OF CONTENTS

INTRODUCTION

Development of advanced software systems for launch support has been an ongoing task at Kennedy Space Center for a number of years. One launch support system which has received much attention in recent years is the loading of liquid oxygen LO_2 and liquid hydrogen LH_2 prior to launch. Two examples of software which have focused on this particular system is the Knowledge-based Autonomous Test Engineer (KATE) developed by Boeing/INET personnel and the Propulsion Advisory Tool (PAT) developed by Rockwell. The loading of LO_2 and LH_2 are very complex physical systems which have several phases of operation. In addition instrumentation aboard the orbiter and ET is limited due to weight constraints. This makes the task of developing software to analyze these systems for fault detection and diagnosis difficult.

The focus of the ten week fellowship has been to demonstrate the development of a robust model for prediction of physical measurements associated with LO_2 loading. More specifically the model was focused on the flow of LO_2 through the orbiter and into the external tank. The development of the prototype model was accomplished with the block diagram simulation language known as VisSim. This report outlines the development effort and presents the results of several simulations. A discussion of how this model can be enhanced and ultimately integrated into existing software such as KATE and PAT is also given.

THE PHYSICAL PROCESS

As previously mentioned this work focused on the loading of liquid oxygen into the external tank of the space shuttle prior to launch. Figure 1 gives a process flow and instrumentation diagram for the LO_2 loading system. Because mass aboard the shuttle must be minimized in order to maximize the payload transported into orbit, instrumentation which may be desirable on the LO_2 system for diagnostics (e.g. flow sensors) is limited. This makes the accurate prediction of system variables such as pressure, temperature and flow desirable.

In order to accurately model the physical system the variability in the operating regimes must be accounted for. There are several stages associated with normal LO_2 loading and additional abnormal conditions. The possible operating conditions for LO_2 and typical count down times associated with them are as follows:

Pump Suction Line Chilldown (T-8:00 to T-6:30)
LO_2 Transfer Line Chilldown (T-6:30 to T-5:50)
LO_2 Orbiter MPS Chilldown (T-5:50 to T-5:30)
Slow Fill ET to 2% (T-5:30 to T-5:15)
Fast Fill ET to 98% (T-5:15 to T-3:25)
Topping of ET to 100% (T-3:25 to T-3:20)
Auto Replenish (T-3:20 to Prelaunch)
Stop Flow (Abnormal)
Revert Flow (Abnormal)

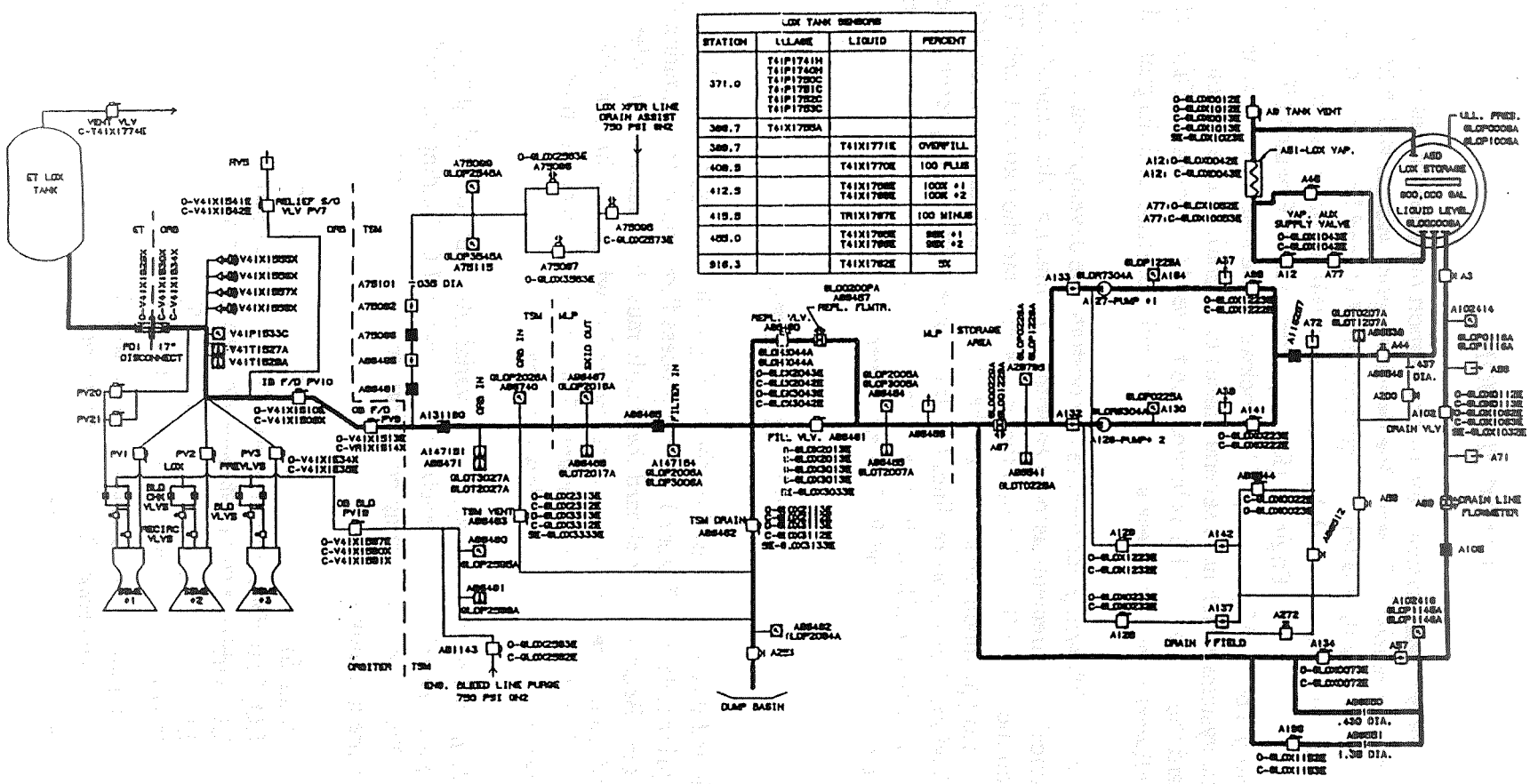


Figure 1 LO_2 Loading Process Flow & Instrumentation Diagram

The chill down periods which are used to condition the transfer system to the liquid cryogen, are the most difficult of all phases in terms of model development. As the cold LO_2 comes in contact with the warm transfer lines and other components, vaporization occurs and two phase flow is encountered. As the LO_2 loading proceeds, the process approaches a thermal steady state and the flow through the transfer system is only in the liquid state. Once the External Tank (ET) is filled, only the mass of LO_2 vaporized from the heat gained through the walls of the tank must be replenished.

As with any physical system which involves the transport of mass and energy, the principals of conservation apply. A conservation of mass equation for the LO_2 ET can be written as:

$$\rho A \frac{dh}{dt} = \rho F_{17''inlet} - \rho F_{ETvent}$$

where :
ρ = *the density of O2*
A = *the cross sectional area of the ET*
h = *the level of LO2*
$F_{17'' inlet}$ = *the flow of LO2 through the 17" inlet*
F_{ETvent} = *the flow of gaseous O2 out the ET vent*

(1)

Due to the geometry of the LO_2 ET the cross sectional area changes as a function of liquid height and thus must be accounted for. For the conditions modeled (i.e. fast fill through replenish) the liquid density of the O_2 can be assumed constant. In general this is not the case however and hence changes in liquid density would need to be accounted for during non isothermal loading phases (e.g. chill down) The flow of a fluid in a pipe between two points a and b is proportional to the square root of the pressure difference. Thus flow terms in the above equation are defined by the equation:

$$F = \alpha(\Delta P)^{0.5}$$

α = *a flow coefficient (admittance)*
ΔP = *the pressure differential for the section of pipe considerred*

(2)

For the orbiter section of Figure 1 the flow of liquid oxygen through the orbiter during normal operation can be described by the equation:

$$F_{PV9} = F_{PD1} + F_{PV1} + F_{PV2} + F_{PV3} \quad (3)$$

In words this continuity equation states that there is no accumulation of liquid in the pipe (i.e. a full pipe) and that the flow into this section through PV9 equals the total flow out of the section through the engine valves PV1, PV2 and PV3 and through the 17" disconnect valve PD1. This equation assumes that the flow is incompressible which is valid for liquid oxygen.

The above equations are the fundamental basis for the development of the predictive models. The procedure of model development was based heavily on loading data from shuttle launches STS50 and STS55. Both of these flights used the Columbia orbiter and in addition used the same launch pad and mobile launch platform. This was desirable from a model building standpoint in order to compare the model against two independent launches while reducing the uncertainty in the physical system.

It should be noted that the system model differs from the actual process in that the engine bleed flows are assumed to branch out from the orbiter inlet as opposed to in the manifold where they actually branch. Although the manifold pressure was calculated from the differential pressure between the ullage and manifold pressures (discussed below) and evaluated for flow prediction, it was not used in the final model. This is due to the fact that the orbiter inlet to 17" disconnect pressure drops gave better model predictions for the flow F_{PD1} given in equation 2.

THE SIMULATION TOOL

VisSim is a block diagram simulation language which can be used to solve both differential and algebraic equations. Modeling equations which describe a physical system are transformed into a block diagram which is numerically solved as a function of time. A graphical interface is provided with pull down menus which allow blocks to be selected and wired together with the use of a mouse. This greatly aides the model building effort in that no program compilation and code debugging is required. In addition, the hierarchal design of the system allows changes to the system model to be readily accomplished.

Other features of VisSim include the use of predefined functions (blocks) which can access other functions or procedures written in C or FORTRAN through user blocks. It also has the capability for Dynamic Data Exchange (DDE) with other applications through DDE blocks. In addition a separate C-code generator and a real time data acquisition package are available as an add on package. Price information for both PC versions and UNIX versions are provided in Appendix A.

MODEL DEVELOPMENT

As previously stated the model building procedure was based heavily on recorded sensor data. Some effort was required to obtain data in the form required by VisSim. Historical flight data which has been recorded and stored on compact disk is readily available. The PC GOAL program which was developed at KSC was used to play back this recorded data and broadcast it over a network. In order to use the data, it first had to be conditioned by recording it to an ASCII file using KATE in the form provided by PCGOAL. A PERL program was then run which could take the recorded file and generate a new ASCII file which could then be read directly into VisSim.

Figure 2 gives the top level of the hierarchal model developed for this work. Each of the blocks given in the Figure represents a combination of additional blocks which describe the individual top level component. The wire connectors and arrows simply represent the flow of information being

passed between individual blocks. To examine any of the blocks at the top level presented in Figure 2, the mouse cursor is moved to that block and a double click of the left button is performed. At any level additional "compound blocks" may be present and again accessed by the same procedure. Each of the blocks at the top level will be discussed below along with how it fits into the model building development.

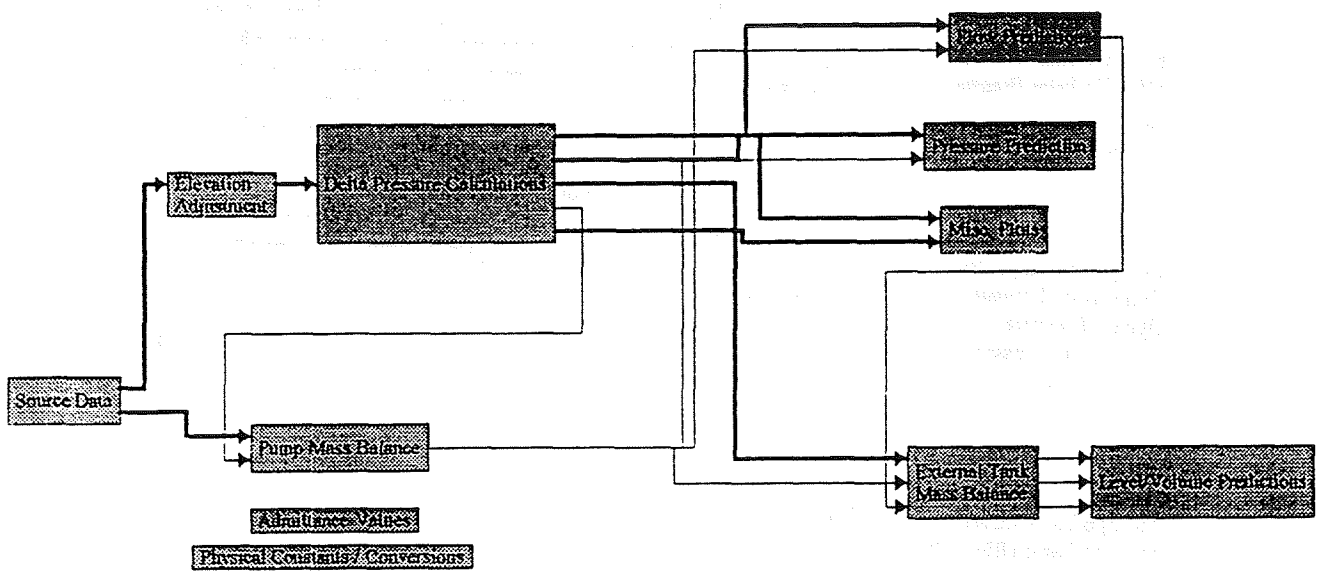


Figure 2 Top Level Block Diagram

The Source Data Block

This block as the name indicates provides the program access to the recorded flight data. Figure 3 gives the block structure for the STS50 data. The data blocks are considered signal providers in VisSim nomenclature and the blocks at this level contain the names of the data files being accessed. The ullage pressure block is simply another compound block which averages the 4 ullage pressure readings from another data file and provides the output back to the top level. Each of the sensor values which are being read from the files is also given and is simply a label in VisSim nomenclature. Also included in this level as well as other level is a scalar to vector block which bundles all the individual wires (scalars) into a single wire (vector).

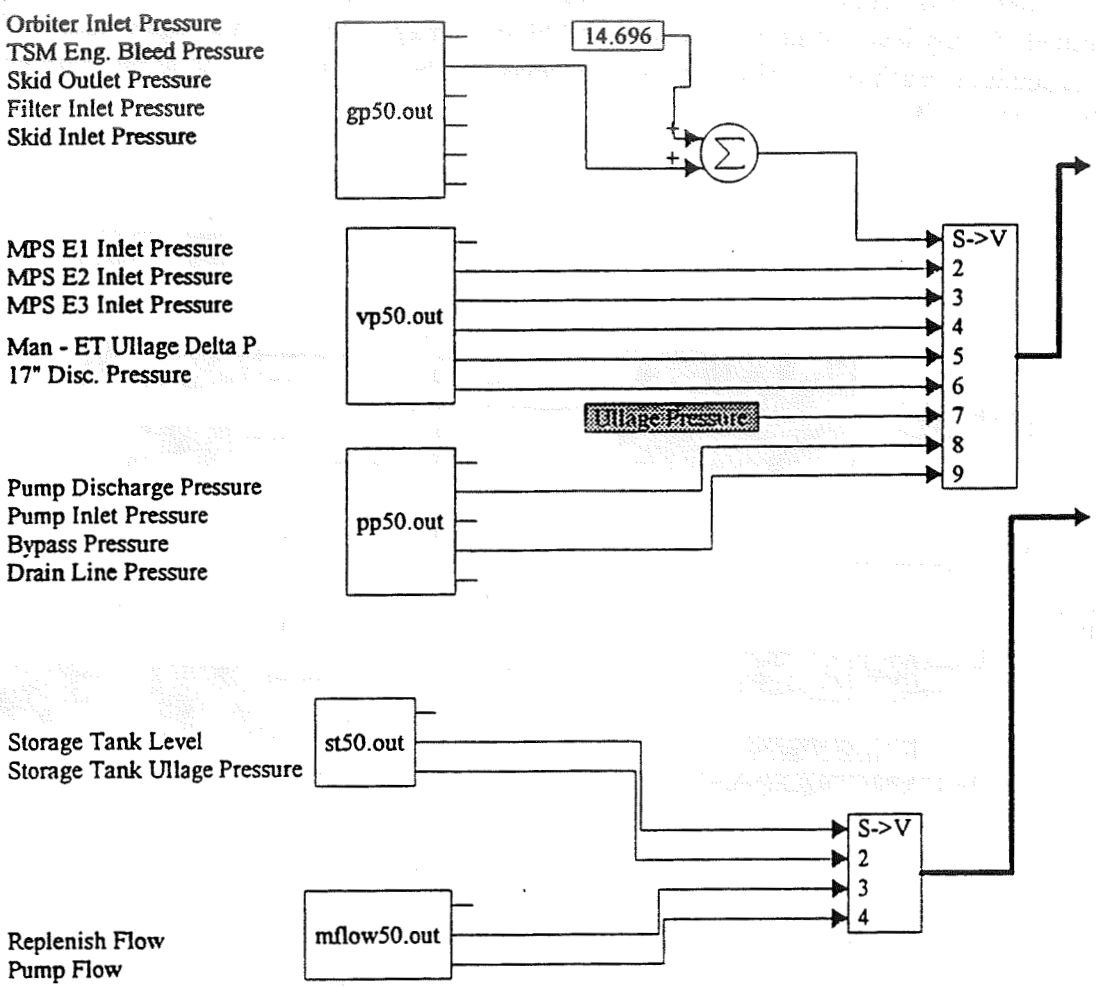


Figure 3 Source Data Block Diagram

The Elevation Adjustment Block

This block which is shown in Figure 4 compensates for the fact that the pressure sensors are located at different elevations. For a static fluid standing in a vertical pipe, the pressure at the bottom of the pipe will be greater than that at the top of the pipe due to the force exerted by the fluid itself. Hence to use the relationship given in equation 2, the elevation difference must be compensated for. The 17" disconnect elevation is used as the zero point to account for elevation differences in the orbiter section. The values given in the constant blocks (5.843, 3.746, etc) which are the elevation corrections in terms of psia (liquid density of 71 lb/ft^3), are subtracted form the individual sensor values. The wires are both unbundled with a vector to scalar block and rebundled after the correction with a scalar to vector block.

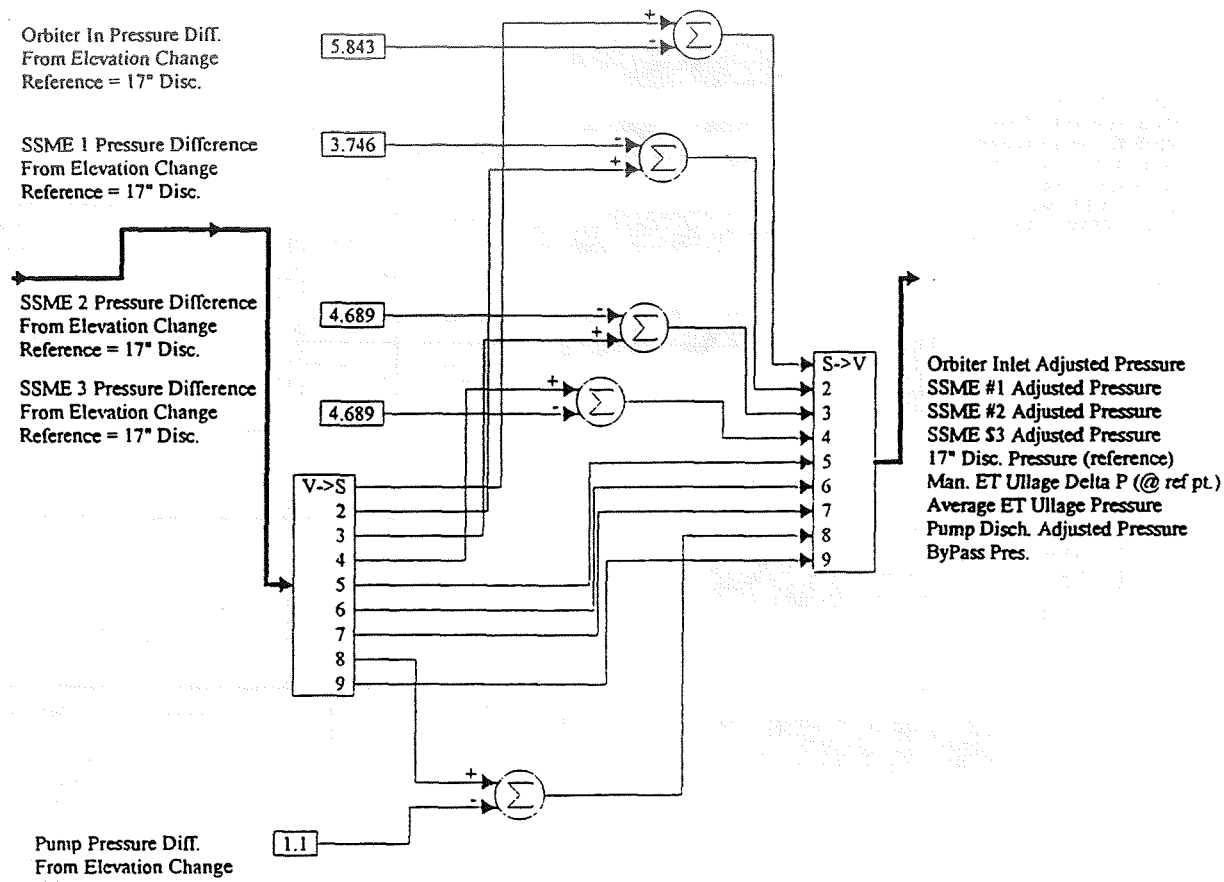


Figure 4 Elevation Adjustment Block Diagram

Delta Pressure Calculations Block

This block which is given in Figure 5 calculates the differential pressures between points of interest (i.e. those in which flow is to be determined).There are 6 separate compound blocks within this block. A typical subblock within this level is given in Figure 6. In this subblock the differential pressure between the orbiter inlet and 17" disconnect is calculated and filtered with a low pass filter. Filters are used in many of the blocks and subblocks throughout the model to reduce the noise in the input data. The variable filter_start is defined under the top level Physical Constants / Conversions block and when used in conjunction with the merge block allows the filtered differential pressure to be passed through whenever the simulation time is greater than 50. Initially however, the unfiltered data is passed through. Using this logic avoids the detrimental dynamic effects associated with filtering if there are no previous values to start from in the input data. (i.e. the initial pressure would be set to zero instead of the measured pressure) The block Dsqrt, which is contained within the subblock, is also used throughout the program and insures that a square root of a negative number is not taken by taking the absolute value prior to the transformation. If the inlet pressure were to be less than the outlet pressure, the flow would be reversed and a negative value would be returned from the block.

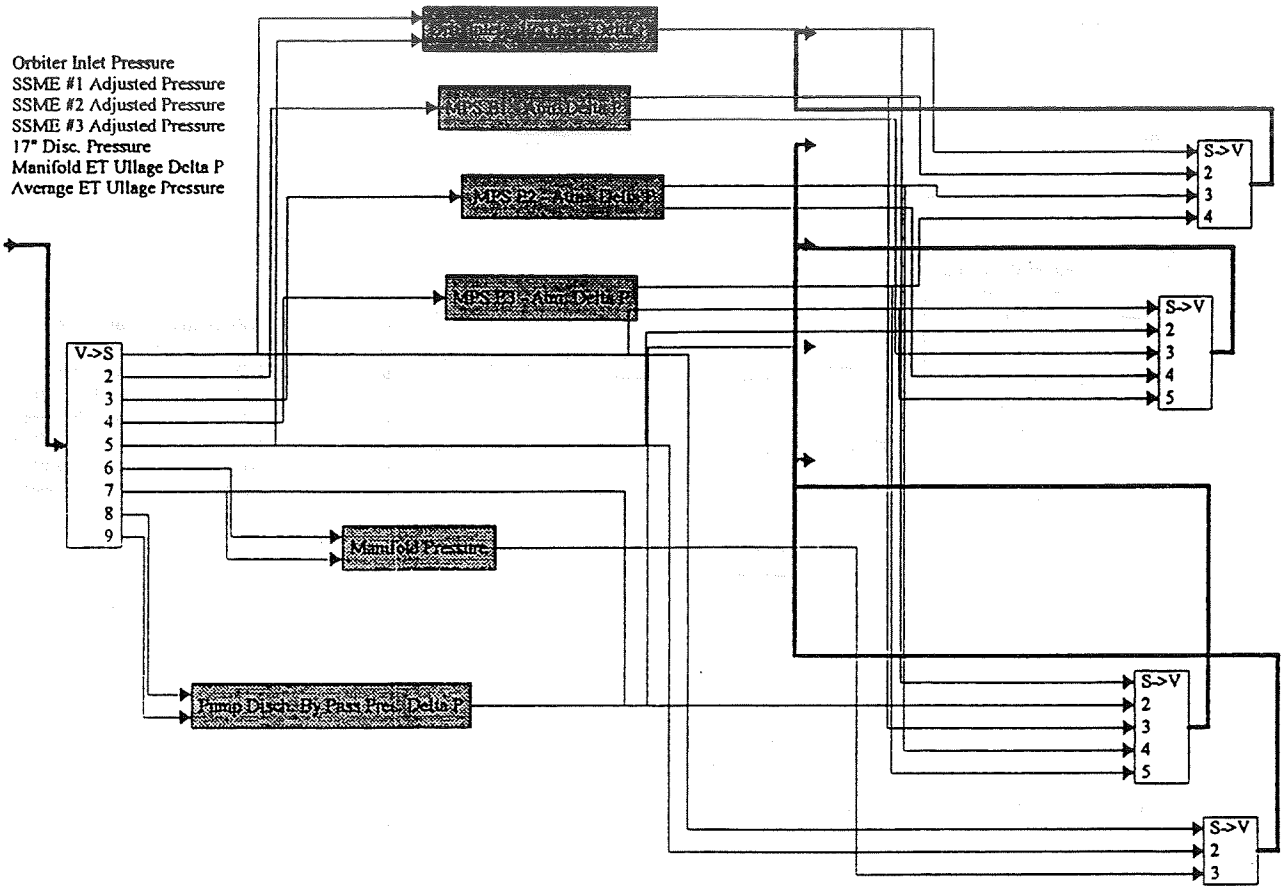


Figure 5 Delta Pressure Calculations Block Diagram

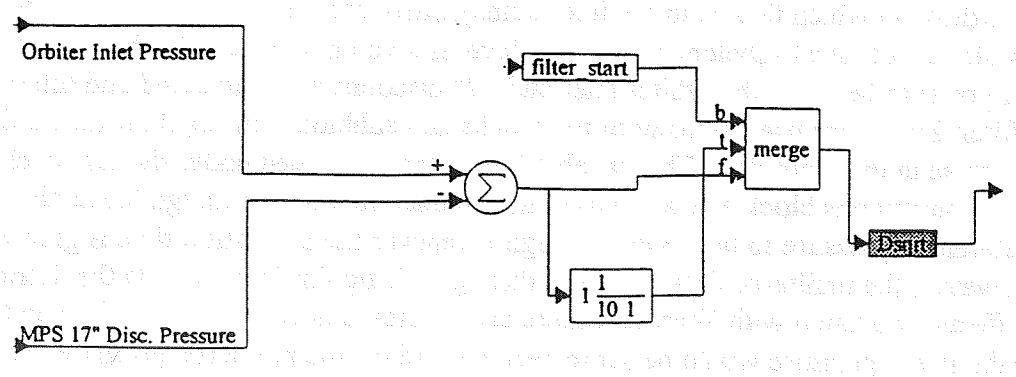


Figure 6 Orbiter Inlet - 17# Disc. Delta P Block Diagram

The other subblocks in Figure 5 consist of one block for each of the engine bleed differential pressures, the calculated differential between the pump discharge pressure and the storage tank recirculation line and the calculation of the manifold pressure. The manifold pressure block does the opposite of the other blocks since the manifold pressure is a differential pressure measurement between the manifold and the ullage pressure. Here the result is obtained by adding the ullage pressure and the barometric pressure (assumed 14.7 psia) to the delta P. Also within the block is a rough estimation of ET level since the differential pressure between the manifold and ullage is proportional to the level of LO_2.

The engine pressures for SSME #2 and SSME #3 are further corrected for offsets which are introduced into the measurement. These offsets were determined from the loading data so that the corrected inlet pressures for the engines were approximately equal. (i.e. as they should be) The pump discharge to storage circulation is essentially identical to the level given in Figure 6.

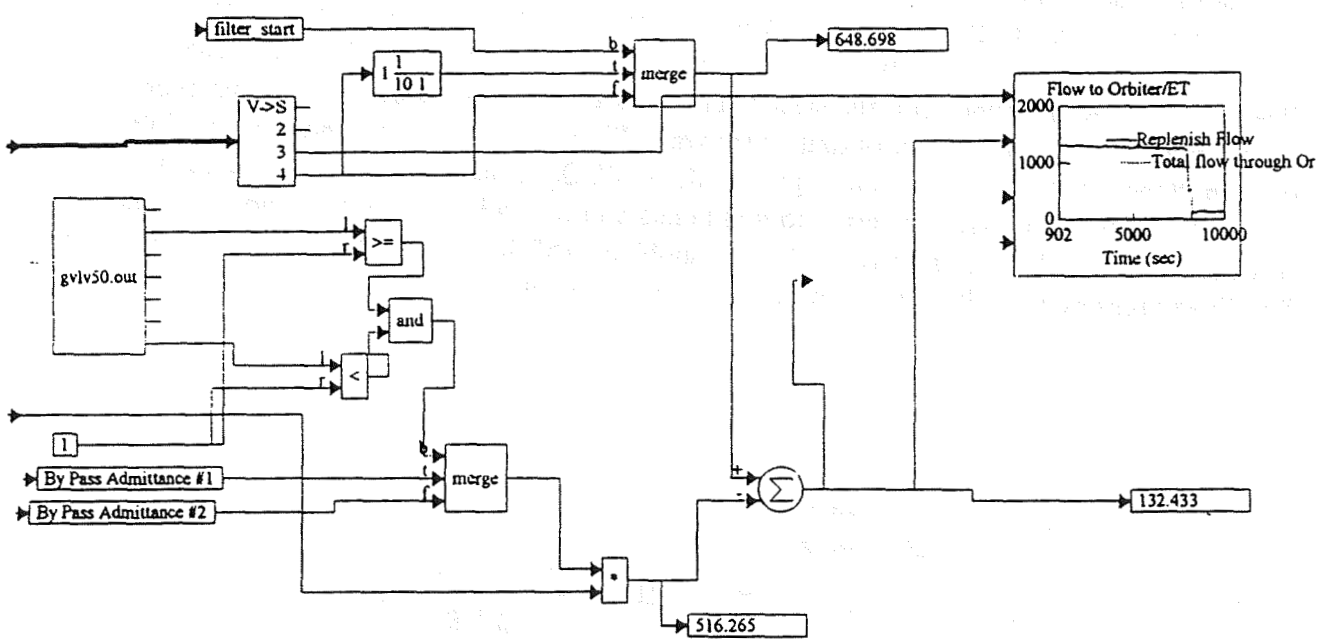


Figure 7 The Pump Mass Balance Block Diagram

The Pump Mass Balance Block

This block which is given in Figure 7 takes the input flow rate as an input and filters it in the same manner that was discussed under the **Delta Pressure Calculations Block**. A mass balance is then performed by subtracting the recycle flow returned to the storage tank. This flow is determined by taking the square root of the differential pressure between the pump pressure and the recirculation pressure calculated from the **Delta Pressure Calculations Block** and multiplying by an admittance value for this line. The admittance can have one of two values depending on whether or not valve A196 (see Figure 1) is opened or closed. As seen in Figure 7, logic is included in the block which examines discrete inputs from a data file (gvl50.out) and when the A196 valve is opened (during topping) a higher admittance value is used to reflect the reduction in resistance in the recirculating line. A plot is also included in this level showing the total flow going to the orbiter/ET once the by pass flow is subtracted. Also included on the plot is the replenish flow. During the replenish model these flows should be equal and this fact was used to determine the admittance value for the recirculating line.

External Tank Mass Balance Block

This block which is presented in Figure 8 contains the core of the model prediction calculations. The flow rate of LO_2 to the ET as calculated from the mass balance discussed in the aforementioned block after subtracting the flow to the engine bleeds (performed in the **Flow Predictions block**) is input into the system The square root of the pressure drop between the orbiter inlet and 17" disconnect output as shown in Figure 6 is also input and then multiplied by the admittance for this section to also get the flow of LO_2 to the ET. Both of these quantities which should be equal, are converted to mass units and then adjusted for the amount of mass flowing from the ET vent valve (if any). The subblock **#Moles / ET ullage Volume** uses the data for ullage pressure and ullage temperature along with the ideal gas law:

$$\frac{n}{V} = \frac{RT}{P}$$

where :

n = *lb-moles of oxygen*

T = *temperature (Rankine)*

P = *pressure (psia)*

R = *gas constant* $\left(10.73 \ \frac{lb\text{-}mole \ \ psia}{ft^3 \ R}\right)$

to calculate the molar volume of gaseous oxygen. The mass flow of gaseous oxygen from the vent valve is then determined by multiplying the output from the subblock **#moles/ET Ullage Volume** by the molecular weight of oxygen (32) and the volumetric flow rate of gaseous oxygen. (determined by the Et vent valve admittance and the square root of the etvent to atmospheric pressure drop)

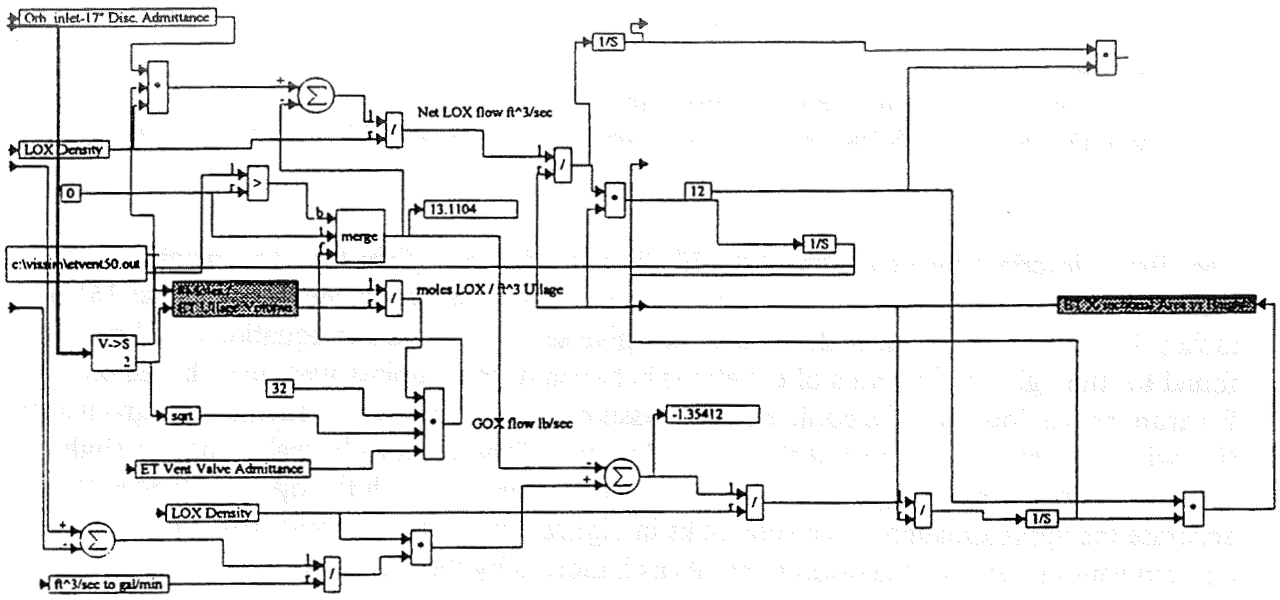


Figure 8 External Tank Mass Balance Block Diagram

The net LO_2 flow into the ET is then determined by dividing by the density of LO_2 and the cross sectional area of the ET. (i.e. a rearrangement of equation 1) The resulting value, which represents the derivative dh/dt is then numerically integrated over time to produce the LO_2 level in the ET. An additional integrator provides the cumulative volume of LO_2 in the ET. This value is fed to the subblock **#moles/ET Ullage Volume** and subtracted from the total volume of the tank in that block in order to calculate the ullage volume.

The cross sectional area of the tank is determined in the subblock **ET X sectional Area vs Height.** This subblock consists of a couple of lookup tables using the VisSim Map block. Once the current height is passed to the subblock the approximate area is determined from one these two data tables (depending on the height) as the Map block employs linear interpolation between the known values.

The tables which are contained in the subblock **ET X sectional Area vs Height** are found in the data files etarea1.dat and etarea2.dat. These files were generated from a separate VisSim program as illustrated in Figure 9. The bottom section of the ET is an ellipse and the equation is given in Et documentation as:

$$\frac{x^2}{165.5^2} + \frac{y^2}{124.125^2} = 1 \qquad (5)$$

where :

x = *the radius of the tank at any point* y

y = *the distance below the* 124.125 *level* (*i.e.* y = 124.125 = *tank bottom*)

The ellipse describes the geometry of the ET from the bottom of the tank to a height of 124.125 inches. From this point to 220.855 inches the tank radius is approximately constant at 165.5 inches. From 220.855 to the tank top are the ogive sections. An exact equation could not be found for the ogive and a series of equations between discrete points were used based on a conic frustrum approximation. The conic frustrum assumes a linear change in radius for a given section. (i.e. where the radius is known at the two end points of the section) In reality there is slight curvature over the entire ogive and thus the more sections in which the ogive is divided the more accurate the approximations. The subblocks in Figure 9 perform the conic frustrum approximations between the discrete points as indicated by the subblock titles.

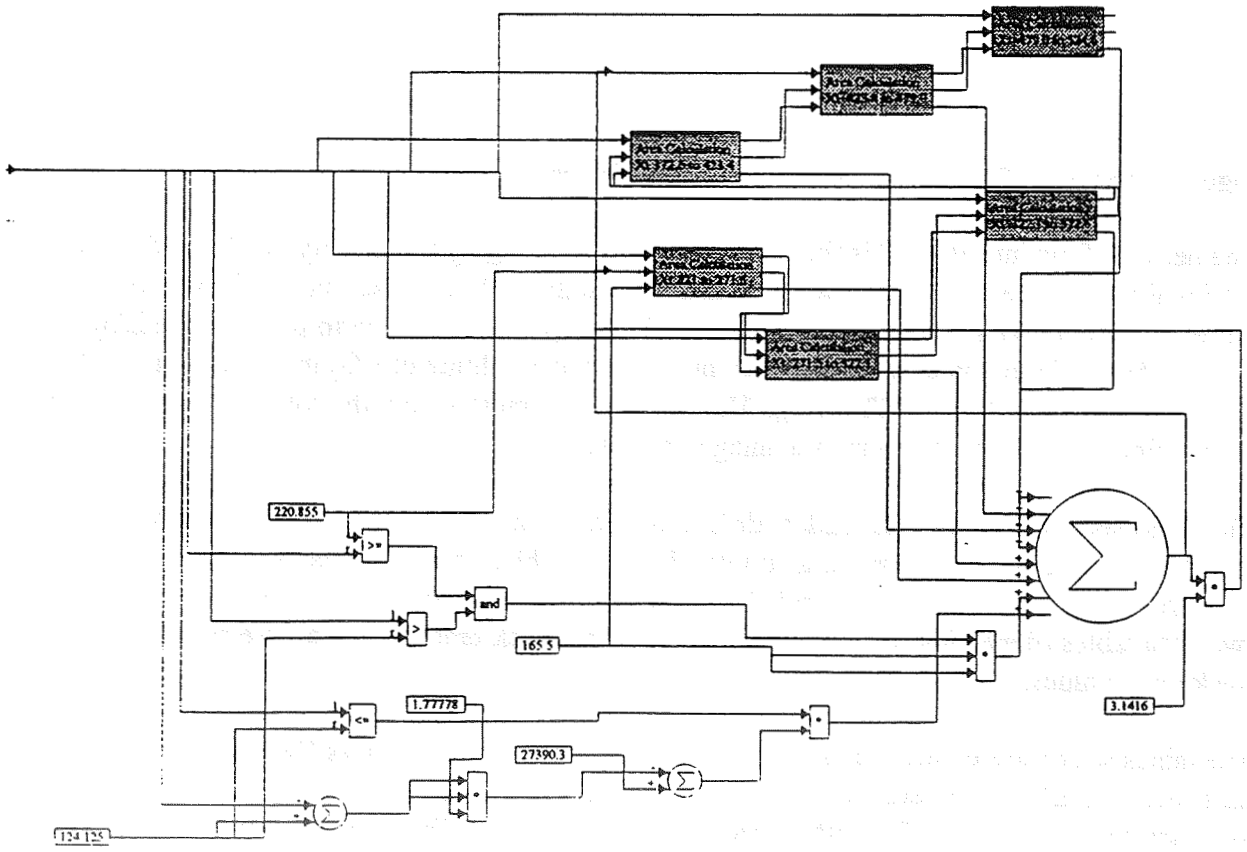


Figure 9 ET X sectional Area vs Height Program Block

Other Blocks

The **Flow Prediction, Pressure Prediction, Misc. Plots and Level/Volume Prediction Blocks** all provide the graphical results of the simulations. These results are provided and discussed in the next section of this report. The **Admittance Values and Physical Constants /Conversions Blocks** are self explanatory.

SIMULATION RESULTS

The primary method for determining unknown admittances in the LO_2 loading system was to iterate on system values until the discrete level sensors in the ET (i.e. 5%, 98%, 100-% and 100%) closely agreed with the actual launch data. For STS50 the predicted level in the tank agreed very closely with actual level at the times when the appropriate sensor levels were reached. For STS55 the predicted level was slightly low relative to the data and indicates that some difference exist between LO_2 loading for these flights.

Figure 10 gives the simulation results of how the LO_2 in the ET changes with time for STS50. The level can be seen to rise continuously until the replenish mode is activated at which time the level is maintained relatively constant. It should be noted that even though there is essentially no change in the level at replenish, the numerical integration is still being performed during this phase and thus indicated that flow in to the ET equals flow out of the ET.

The prediction of the flow rate through the 17" disconnect going to the ET is presented in Figure 11 for flight STS 50. As previously discussed, this flow is predicted by both a mass balance and a pressure drop admittance calculation. The mass balance approach gives the more reliable results and was used to determine the LO_2 level in the ET presented in Figure 10. The flow calculation using the pressure differential is very noisy and provides very erratic results including reverted flow during the replenish phase as evidenced in Figure 11.

The prediction of flow for STS 50 through the engine bleeds is presented in Figure 12. The admittance on these predictions was adjusted so the flows would fall in between the 18 - 20 gpm range which is thought to be a close estimate.

The prediction of corrected pressure values for the orbiter inlet, 17" disconnect and engine pressures for STS 50 is given in Figures 13, 14 and 15 respectively. Figures 16, 17, and 18 are the same pressure predictions for the STS 55 flight using the same admittance values. It can be observed that the actual and predicted values for both flights agree in a similar fashion thus validating the model developed. Figure 19 gives the flow rate through the 17" disconnect going to the ET for STS 55. It can be observed that besides predicting a lower level than actually achieved based on the sensor locations, the level is shown to gradually decrease during replenish, thus reconfirming the need for additional data evaluation.

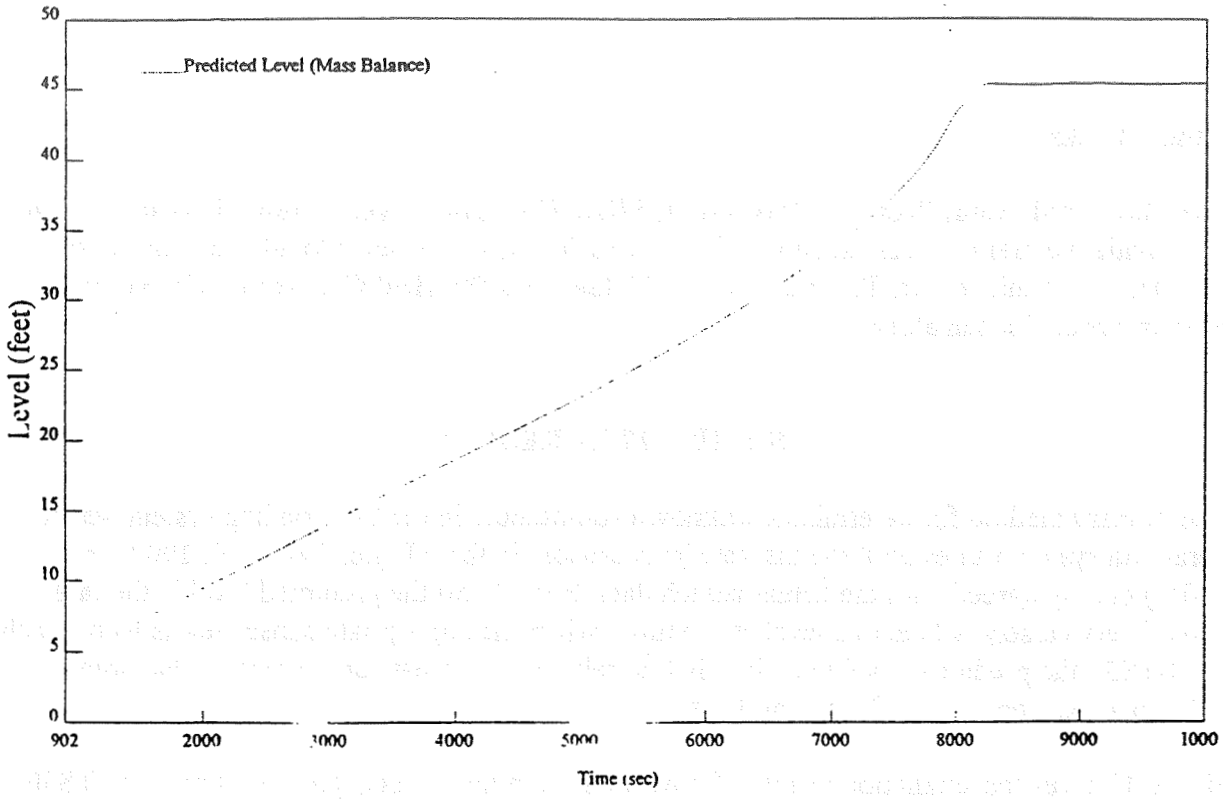


Figure 10 Simulated ET LO_2 Level Vs. Time (STS50)

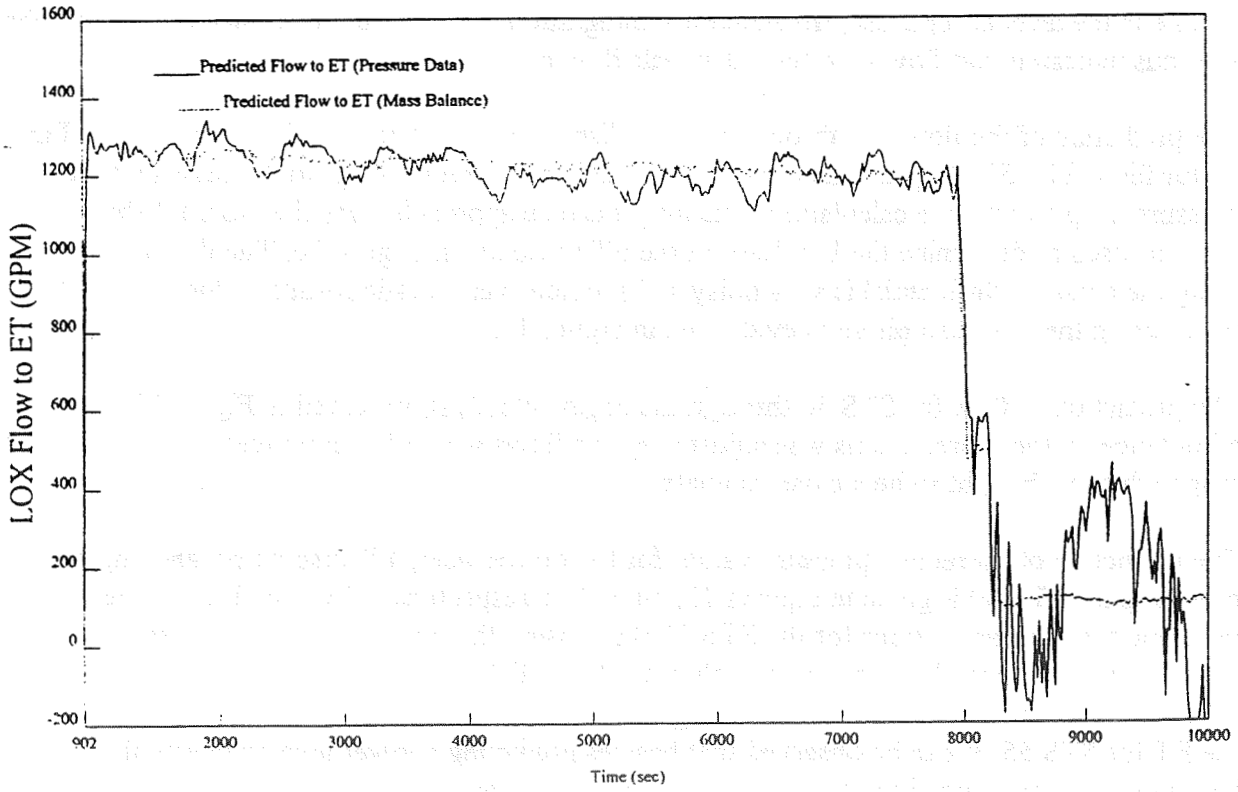


Figure 11 Predicted LO_2 Flow Rates to ET (STS50)

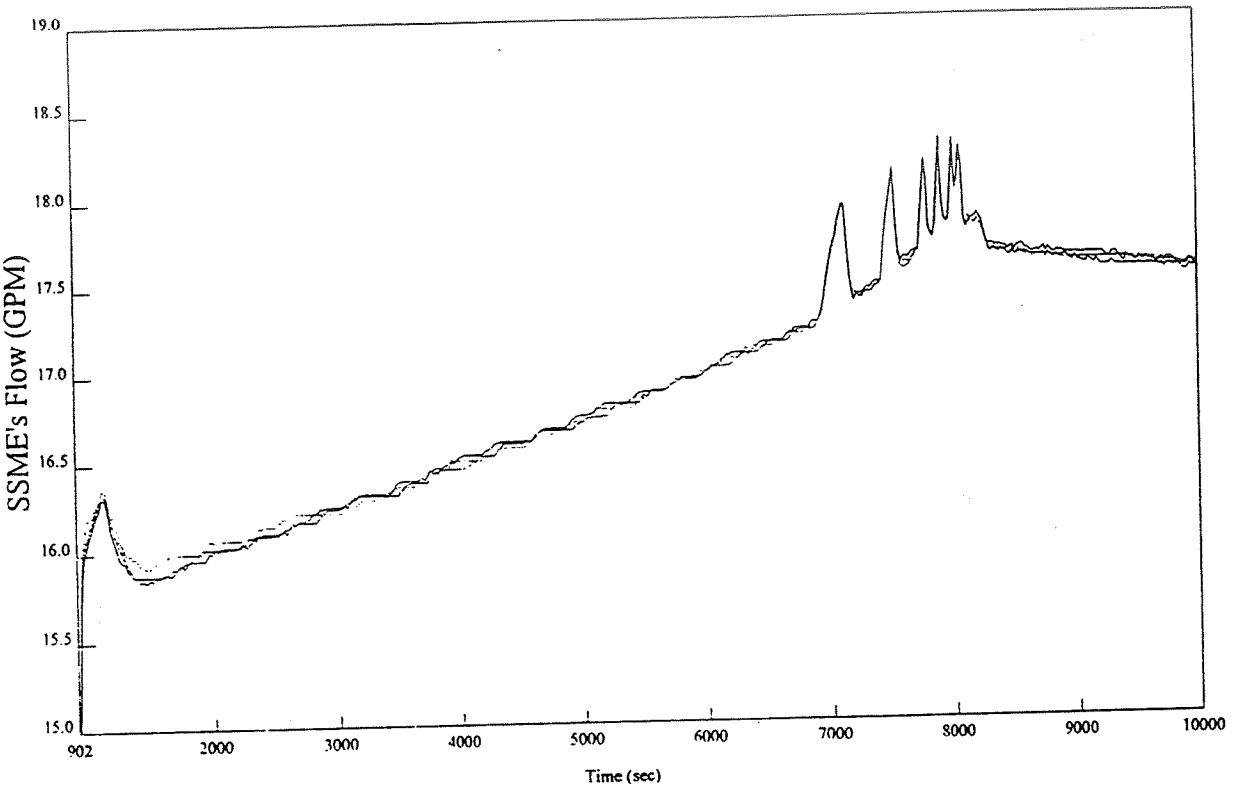


Figure 12 Predicted LO_2 Bleed Flow Rates for SSME's (STS50)

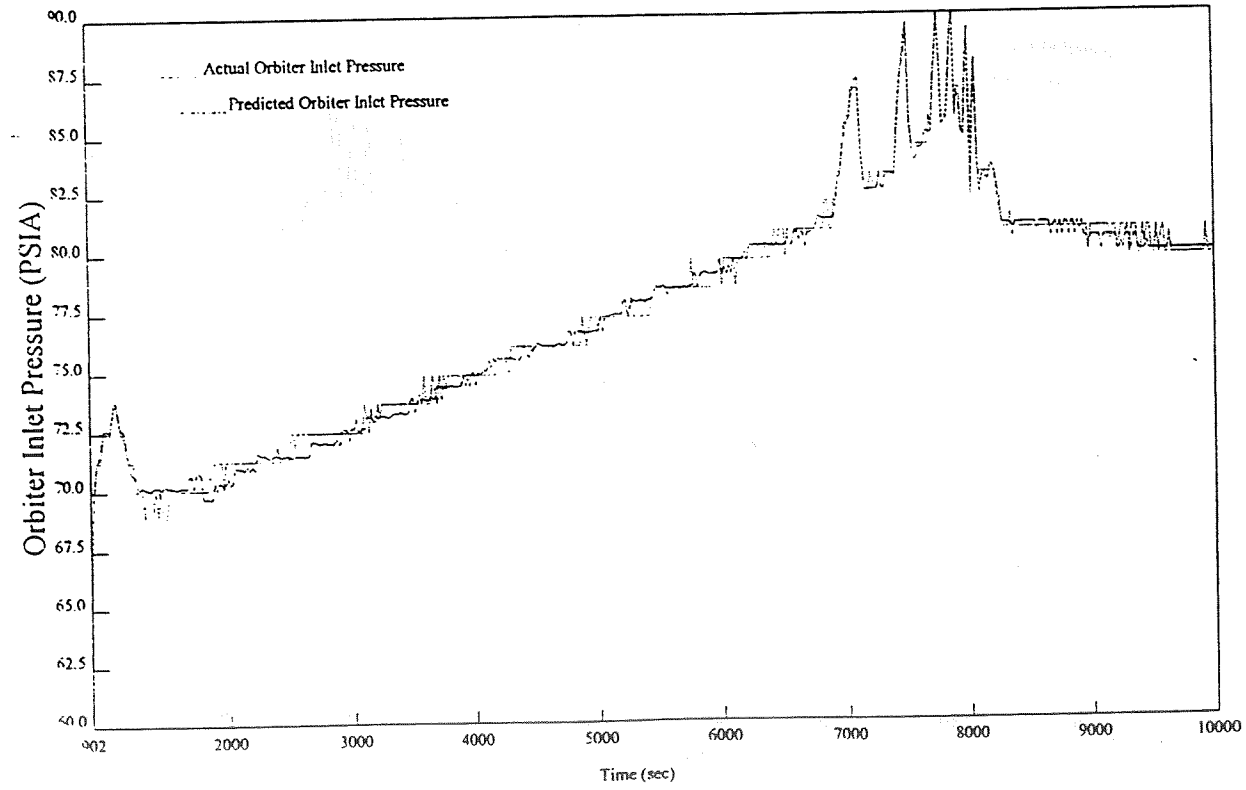


Figure 13 Predicted Vs. Actual Orbiter Inlet Pressure (STS50)

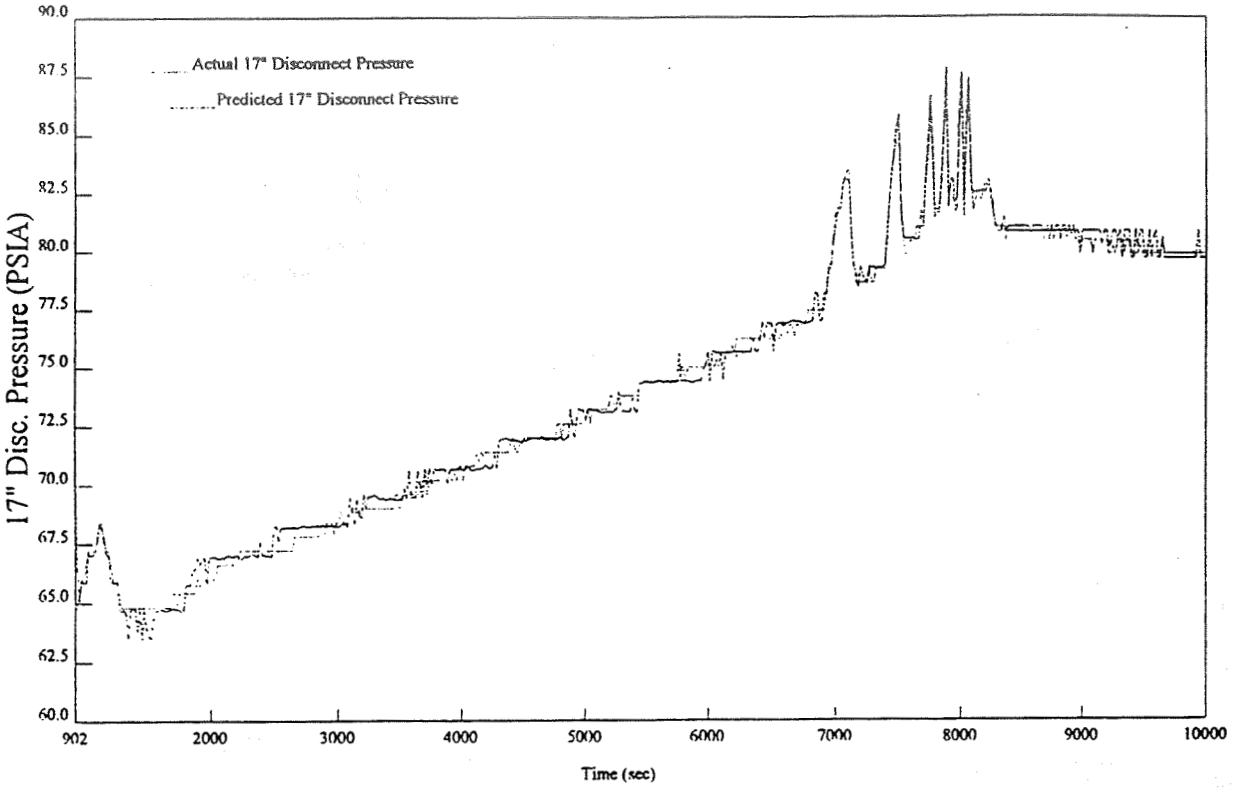


Figure 14 Predicted Vs. Actual 17" Disconnect Pressure (STS50)

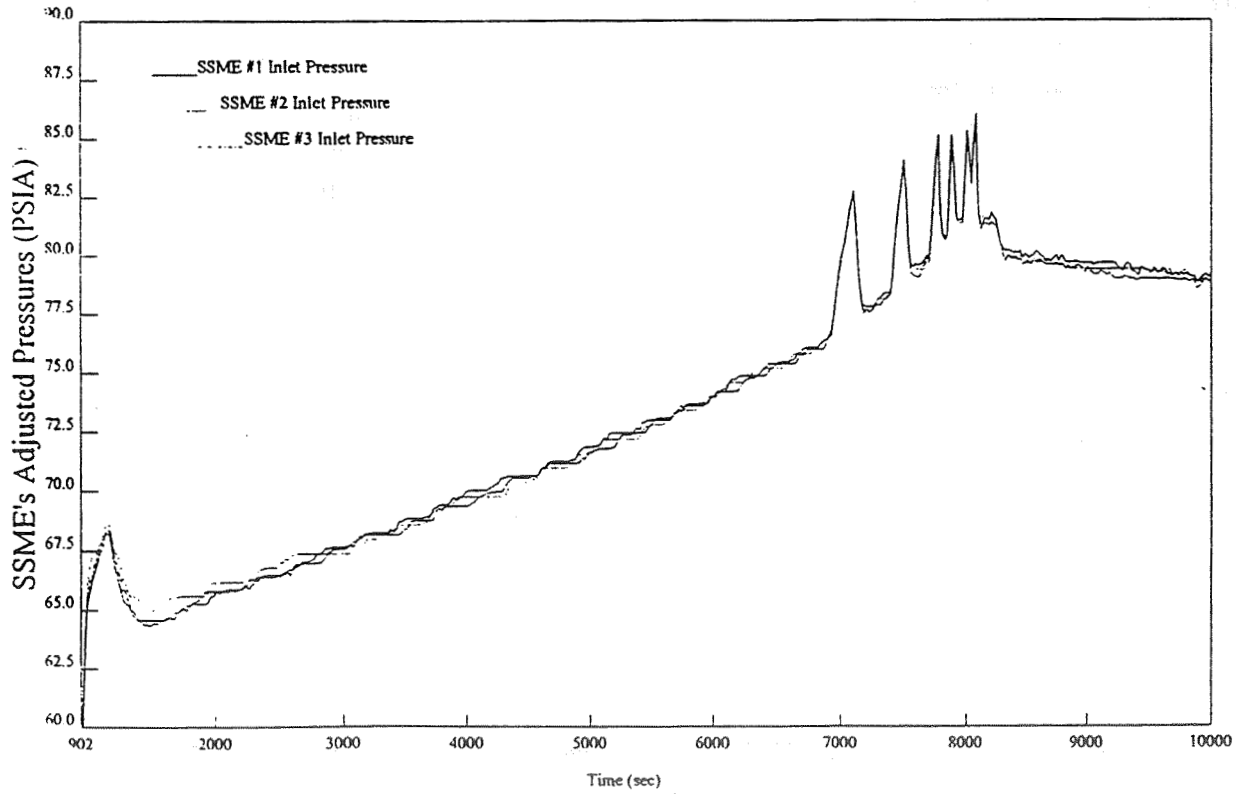


Figure 15 Predicted Vs. Actual SSME Bleed Pressures (STS50)

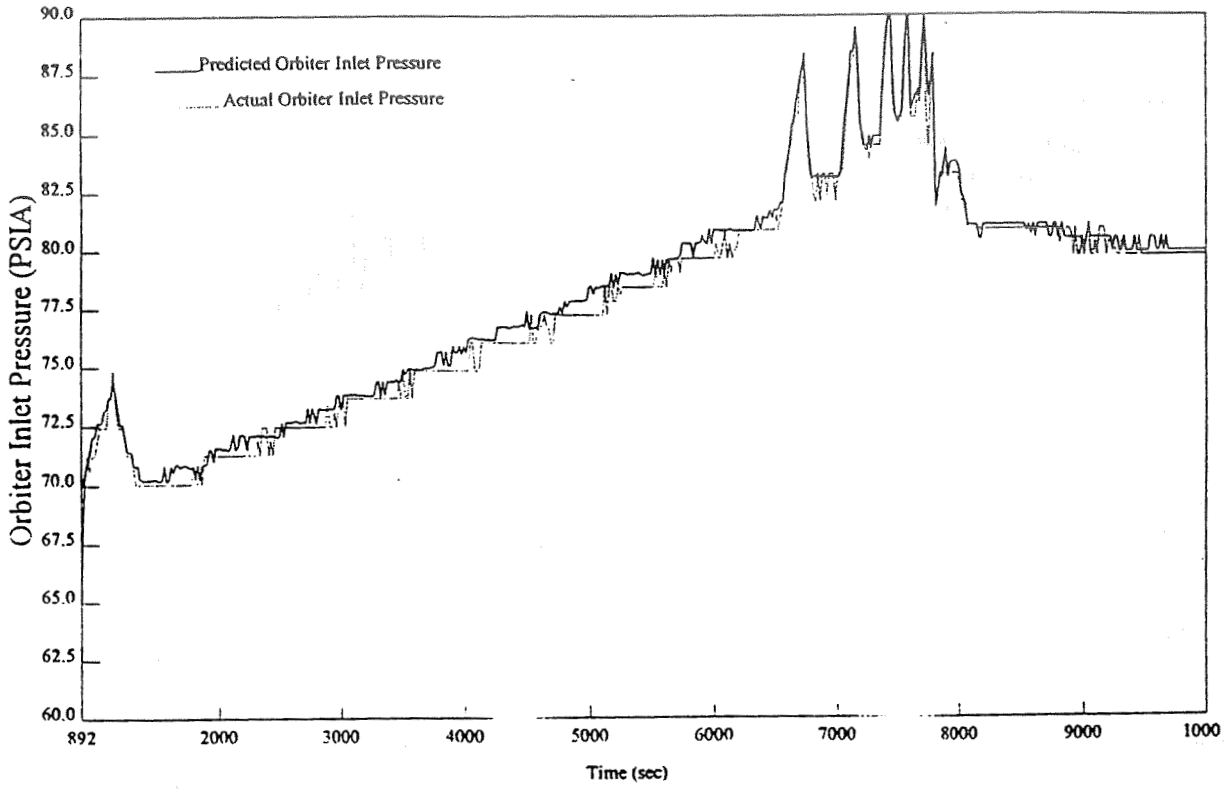


Figure 16 Predicted Vs. Actual Orbiter Inlet Pressure (STS55)

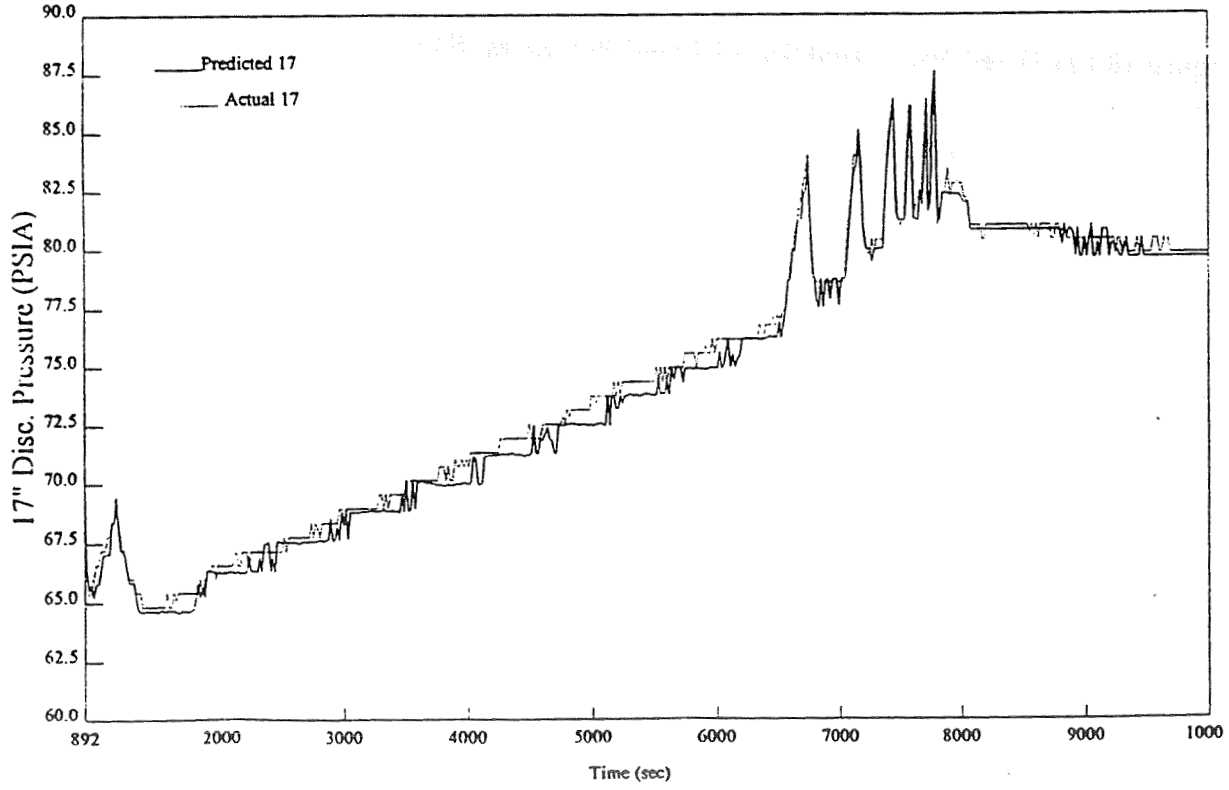


Figure 17 Predicted Vs. Actual 17" Disconnect Pressure (STS55)

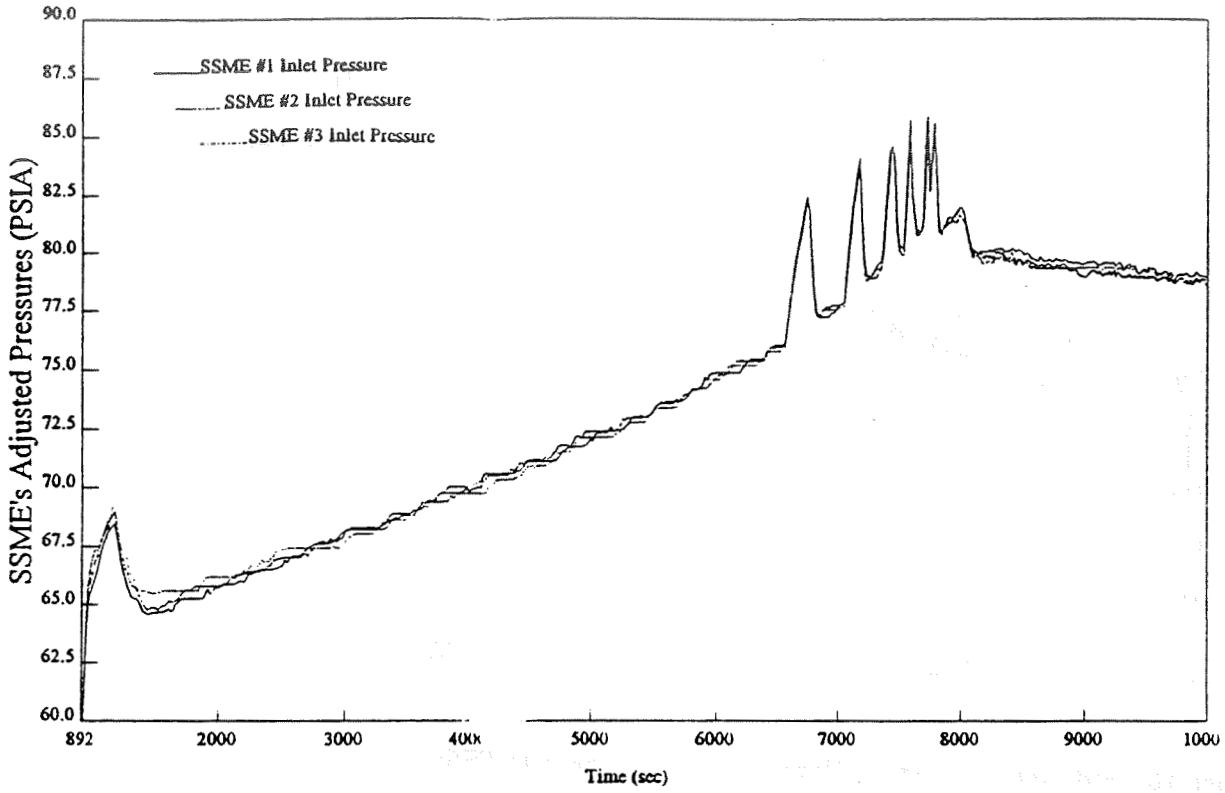


Figure 18 Predicted Vs. Actual SSME Bleed Pressures (STS55)

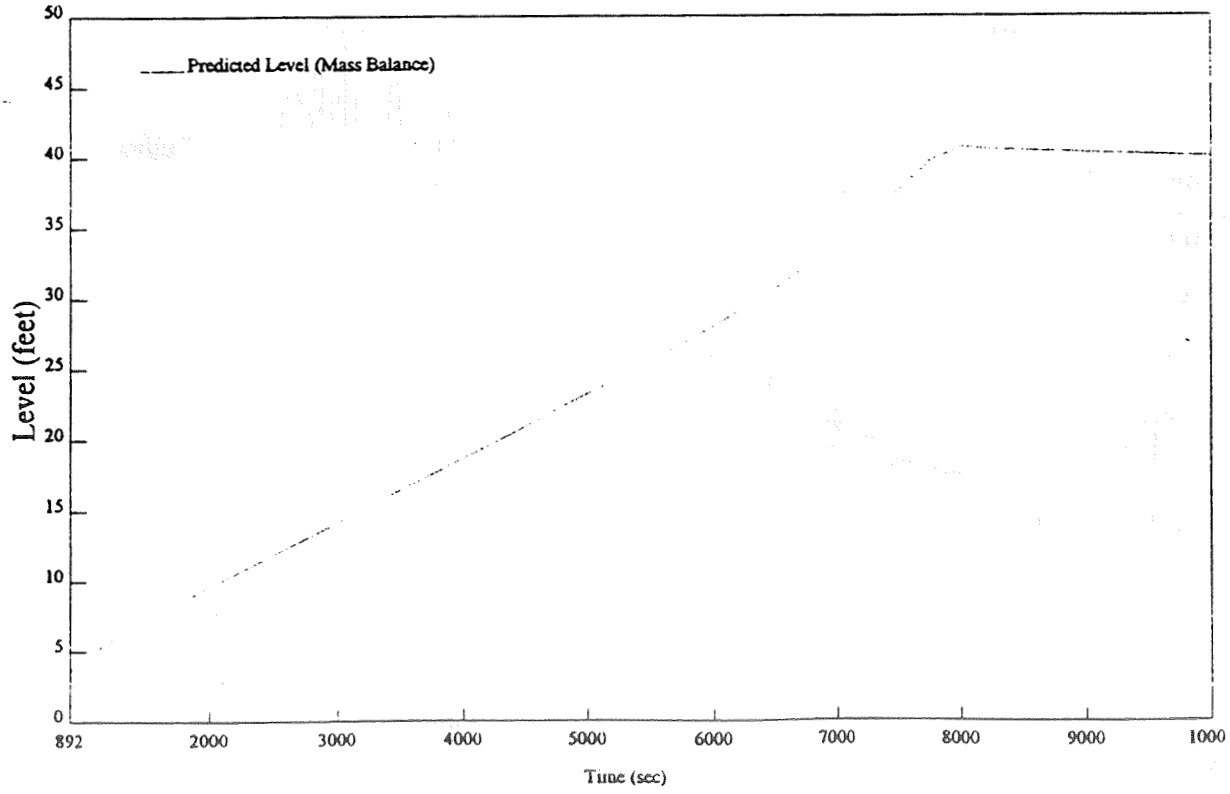


Figure 19 Predicted LO_2 Flow Rates to ET (STS55)

FAULT DETECTION

VisSim can also be used to detect faults within the system by observing the differences between the actual and predicted values and determining whether or not they are within normal tolerances. To confirm that the predicted and actual values would result in discrepancies if a fault entered the process a single point failure was simulated. The results which are shown in Figures 20, 21 and 22 show that the predictive values no longer agree with the actual values at approximately 7300 seconds into the simulation. At this point a false signal was sent to the program that the recirculation valve had opened. As a consequence the admittance was changed on this section of pipe by the program resulting an error in the mass balance when compared to the true process. This error is propagated through the pressure measurements also as evidenced by the plots. When the valve is actually opened at approximately 7650 seconds the predicted and actual measurements agree again.

CONCLUSIONS & RECOMMENDATIONS

The use of VisSim has been shown to be a useful tool for robust and rapid model building of a complex process. The development effort put forth this summer was severely constrained, due to the fact that a personal version of VisSim was employed instead of the upgraded professional version. The personal version is limited in the number of blocks which can be used in the model whereas the professional version has no such constraint. The block limitation became a frequent source of model redesign in order to work within the constraint and thus hampered the development effort.

While the model building effort needs to be extended to cover additional loading regimes and thermal predictions, the current model can serve as the building block for this effort. A professional version of VisSim should be procured before substantial enhancements of the model can be achieved. Additional loadings should also be investigated to determine the effects (if any) of the launch pad, mobile launch platform and orbiter on the predictive models. If differences were encountered as might be expected the model could be adapted (tuned) for each individual launch.

To incorporate the model into an existing software package such as KATE or PAT a mechanism would need to be set up which could link the VisSim application to the existing program. VisSim supports dynamic data exchange and hence may be able to be run in parallel with the existing application. In this manner, pertinent loading data would be passed to the VisSim Environment and the predicted values could either be passed back to the native application or faults could be detected within VisSim and information only passed back when faults occur. If the two processes could not operate in parallel two other options exist; either the model could be encoded into the language native to the particular application or a C-code generator for VisSim could be purchased. The C-code generator would be the more desirable approach as it generates C source code but allows model development to be performed using the graphical block diagram environment.

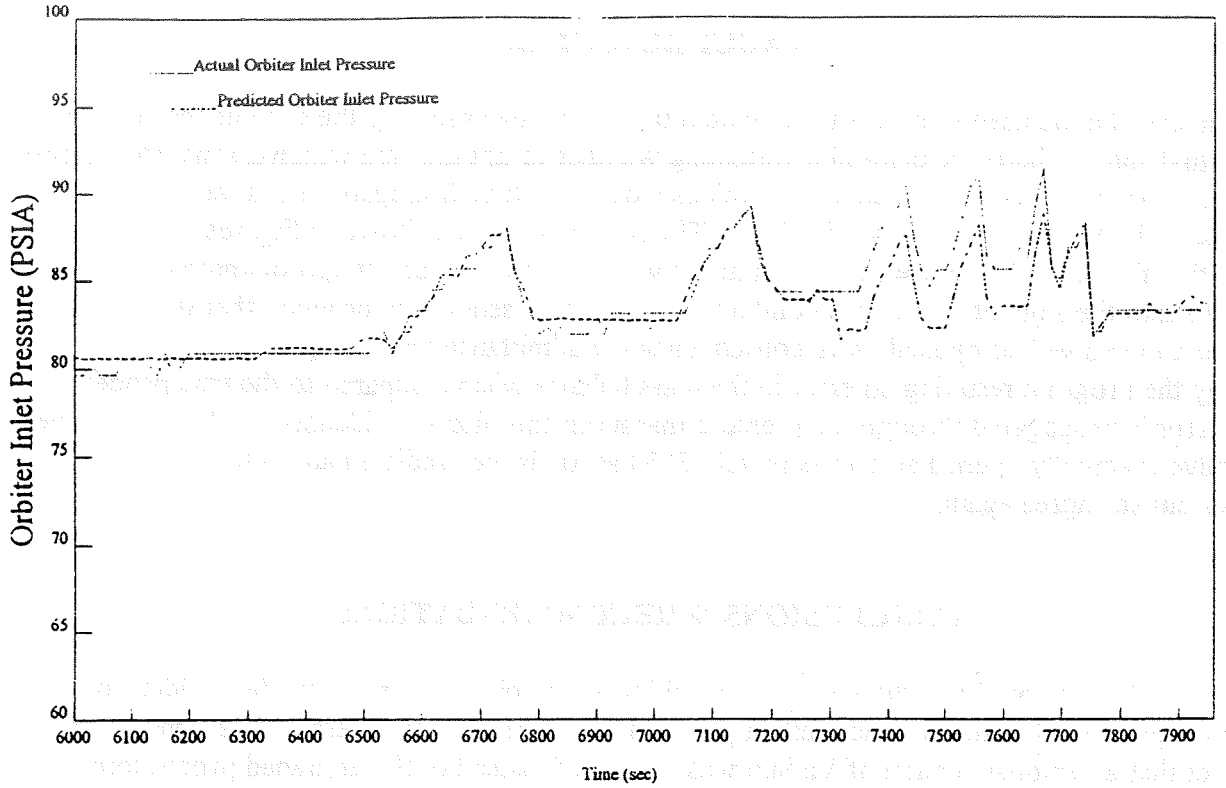


Figure 20 Failure Induced Predictive Error for the Orbiter Inlet Pressure (STS55)

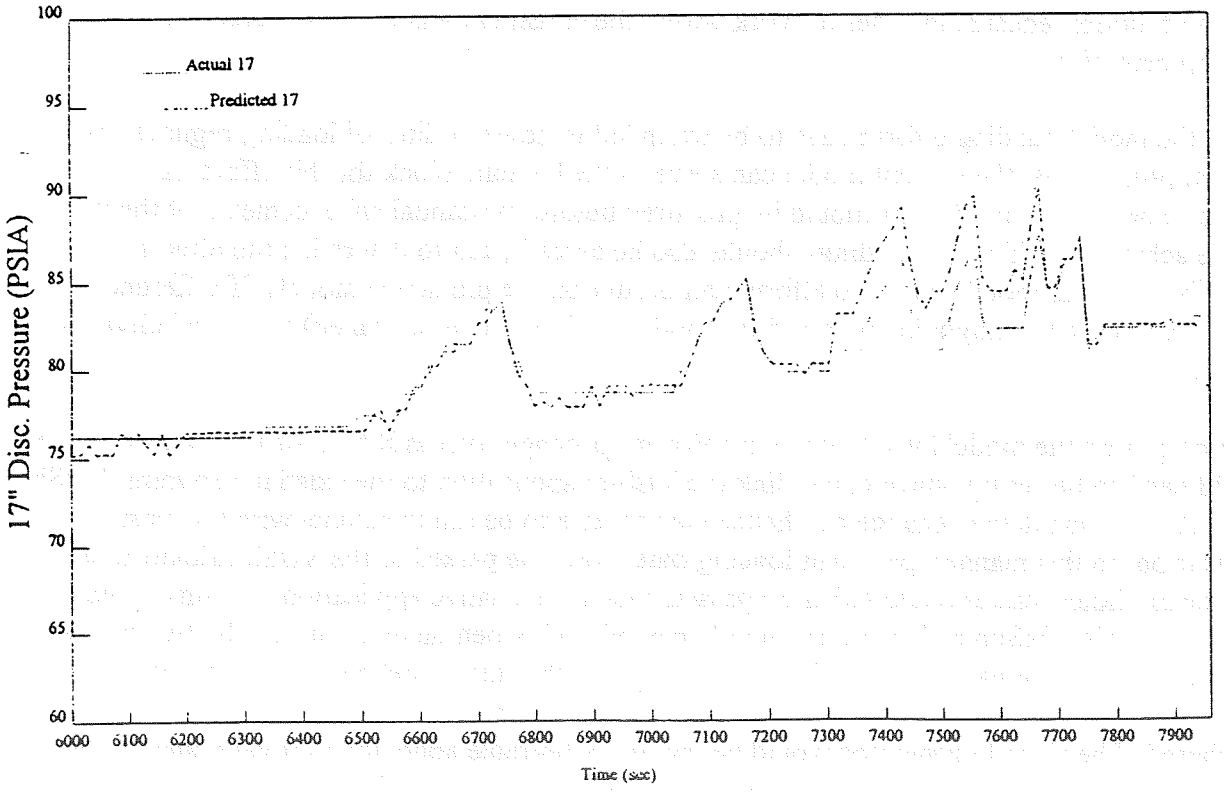


Figure 21 Failure Induced Predictive Error for the 17" Disconnect Pressure (STS55)

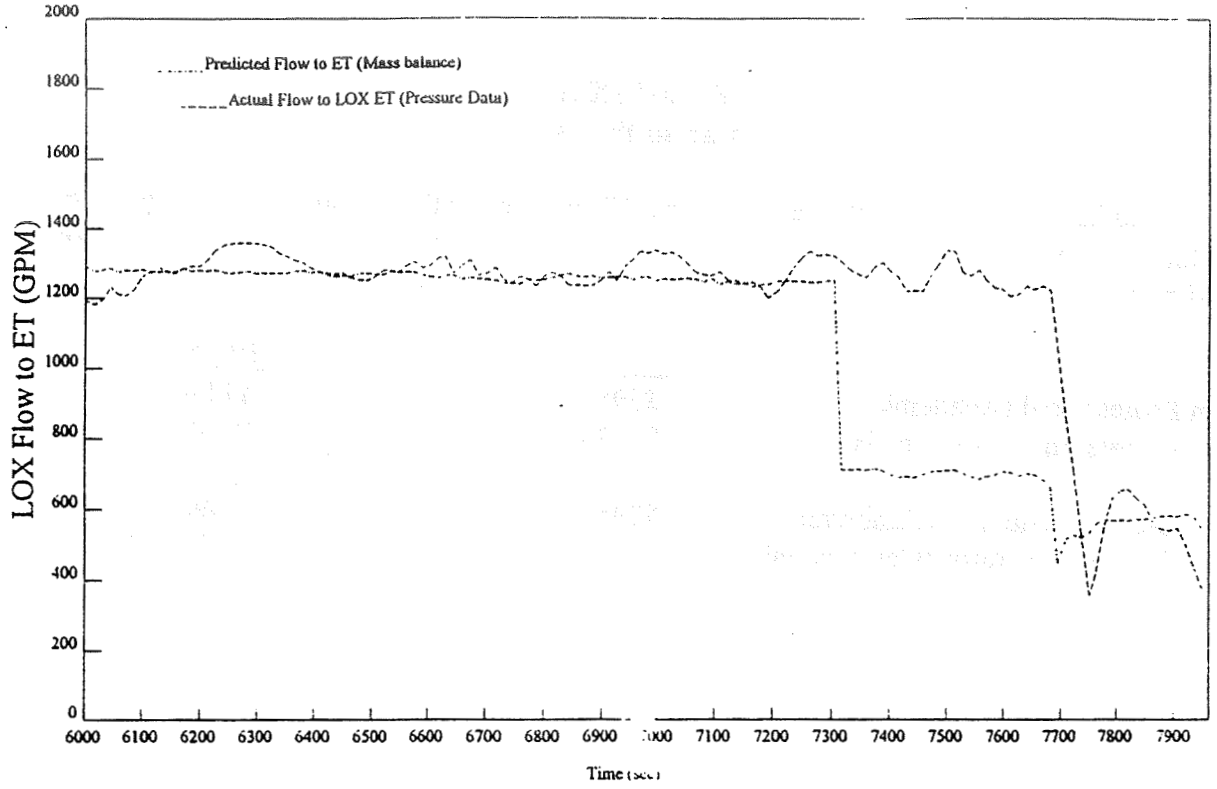


Figure 22 Failure Induced Predictive Error for the LO_2 Flow Rates to ET (STS55)

APPENDIX A
VisSim Prices

Current price information for both the UNIX and PC versions of the VisSim Program and C-code generator is provided below. The academic prices are exact, while the commercial versions are approximate.

	PC	UNIX
VisSim Professional (Academic)	$595	$1190
VisSim Professional (Commercial)	$1795	$2140
VisSim C-code Generator (Academic)	$345	$690
VisSim C-code Generator (Commercial)	$995	$1200

1995 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM
JOHN F. KENNEDY SPACE CENTER
UNIVERSITY OF CENTRAL FLORIDA

S22-61
7762
P-30

FOLLOW-THE-LEADER ALGORITHM FOR THE PAYLOAD INSPECTION AND PROCESSING SYSTEM

Dr. Robert L. Williams II
Assistant Professor
Mechanical Engineering Department
Ohio University
Athens, Ohio

KSC Colleagues - Eduardo Lopez del Castillo and Gabor Tamasi
Robotics

Contract Number NASA-NGT-60002
Supplement 19

August 18, 1995

FOLLOW-THE-LEADER ALGORITHM FOR THE PAYLOAD INSPECTION AND PROCESSING SYSTEM

Robert L. Williams II, Ph.D.

ABSTRACT

This report summarizes the author's summer 1995 work at NASA Kennedy Space Center in the Advanced Systems Division. The assignment was path planning for the Payload Inspection and Processing System (PIPS). PIPS is an automated system, programmed off-line for inspection of Space Shuttle payloads after integration and prior to launch. PIPS features a hyper-redundant 18-dof serpentine truss manipulator capable of snake-like motions to avoid obstacles. The path planning problem was divided into two segments: 1) Determining an obstacle-free trajectory for the inspection camera at the manipulator tip to follow; and 2) Development of a follow-the-leader (FTL) algorithm which ensures whole-arm collision avoidance by forcing ensuing links to follow the same tip trajectory.

The summer 1995 work focused on the FTL algorithm. This report summarizes development, implementation, testing, and graphical demonstration of the FTL algorithm for prototype PIPS hardware. The method and code was developed in a modular manner so the final PIPS hardware may use them with minimal changes. The FTL algorithm was implemented using MATLAB software and demonstrated with a high-fidelity IGRIP model. The author also supported implementation of the algorithm in C++ for hardware control.

The FTL algorithm proved to be successful and robust in graphical simulation. The author intends to return to the project in summer 1996 to implement path planning for PIPS.

ACKNOWLEDGMENTS

The author acknowledges Gabor Tamasi of the Advanced Systems Division at KSC for an interesting theoretical and hardware problem, and for guidance in its solution. Many thanks also to Carey Cooper of Intergraph Corporation who performed the IGRIP modeling and provided enlightening insights.

TABLE OF CONTENTS

FIGURE LIST

TABLE LIST

I. INTRODUCTION

1.1 BACKGROUND

Inspection of Space Shuttle payloads after integration and prior to launch is essential for launch and mission safety. Currently, this inspection is completed by humans, which is dangerous, costly, labor-intensive, and not versatile in the cluttered and sensitive Shuttle bay environment. With shrinking budgets, development of efficient, labor saving methods are warranted. Therefore, the Advanced Systems Division at NASA Kennedy Space Center (KSC) is developing an automated tool, the Payload Inspection & Processing System (PIPS), for prelaunch inspection and light tasks in the Space Shuttle bay [1],[2]. Figure 1 shows the design concept for PIPS. This unique device features a hyper-redundant serpentine truss manipulator (STM) for carrying a camera along obstacle-free trajectories to required goal points for inspection. The prototype PIPS hardware (built by Foster-Miller, Inc. [3] and modified by NASA) has eighteen degrees-of-freedom. The development of PIPS hardware for use at either lauchpad Payload Changeout Room (PCR) is underway. NASA is also interested in serpentine manipulators for in-space construction [4] and in-space inspection tasks [5].

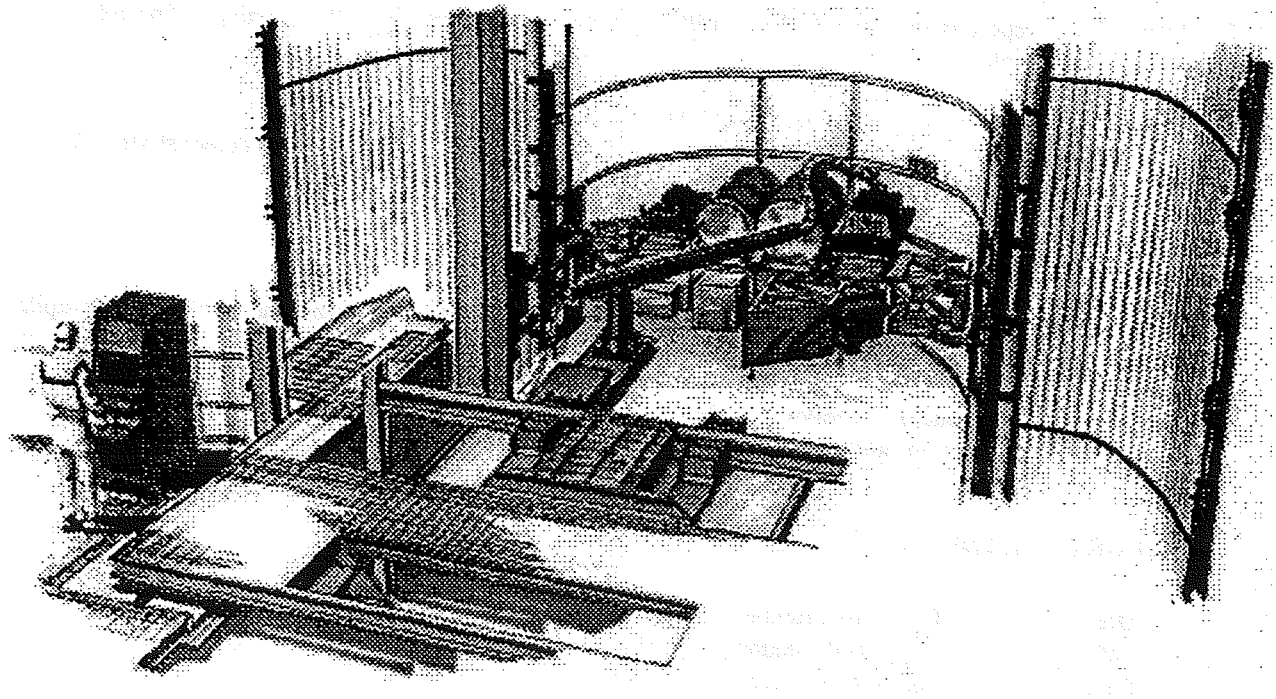


Figure 1.1 PIPS Conceptual Design

1.2 THE PATH PLANNING PROBLEM

PIPS is an automated system which will be programmed off-line. Path planning is critical for PIPS to avoid contact in the cluttered Shuttle payload environment. The robotics literature is rich with path planning and obstacle avoidance techniques (e.g. [6], [7], [8], [9]). Path planning for this project is divided into two phases: 1) Determination of collision-free manipulator tip trajectories to reach required goal points. 2) Development and

implementation of a follow-the-leader (FTL) algorithm which forces ensuing links to follow the same tip trajectory. Given an obstacle-free manipulator tip trajectory, the FTL algorithm ensures whole-arm collision avoidance. The path planning algorithm must be constrained to yield trajectories which are achievable using the FTL algorithm with specific manipulator kinematics and constraints such as joint limits.

The author presents a general, modular FTL algorithm. This is an extension of the methods in [2] and [3]. Asano [6] suggests a similar scheme. The FTL algorithm was derived, implemented, and tested for the prototype hardware. It is adaptable with minimum change to new manipulator kinematics.

The author proposes to return to the PIPS project in summer 1996 to focus on the path planning algorithm. Information gathering and preliminary development for path planning was accomplished in summer 1995. Three methods are recommended, in order of preference: 1) Human-generated obstacle-free trajectories using IGRIP for visualization. 2) Same as 1), with the human producing the initial global trajectory. A computer algorithm then would fine-tune this result subject to potential field and curvature (joint limit) constraints. This mode is termed human-assisted path planning. 3) Automatic computer generation of trajectories, where the manipulator tip link is placed at the goal position and orientation. In-board links are placed optimally in turn, using potential fields .

1.3 REPORT ORGANIZATION

The focus of this report is the PIPS FTL algorithm development for use in path planning. The following are summer 1995 accomplishments:

- FTL algorithm development, MATLAB implementation, and testing
- High-fidelity prototype hardware modeling in IGRIP (with Carey Cooper) to demonstrate FTL
- Kinematics modeling and information compilation for prototype hardware
- Worked with programmer to implement FTL algorithm in C++
- Preliminary development for path planning algorithm

This report is organized as follows. Section II presents kinematic modeling and information compilation for the prototype PIPS hardware, plus a discussion of the developed IGRIP model. Section III presents the general FTL algorithm. The subsections give detailed derivations of the required mathematics. Section III concludes with simulated results. The conclusion follows, including a summary of design lessons learned from simulation. The last section is references. Four appendices are given to support the text.

1.4 REPORT ACRONYMS

DH	Denavit-Hartenberg
dof	degrees-of-freedom
FTL	Follow-the-Leader
IGRIP	Commercial Robot Simulation Software
KSC	Kennedy Space Center
MATLAB	Commercial Control Simulation Software
NASA	National Aeronautics and Space Administration
PCR	Payload Changeout Room
PIPS	Payload Inspection and Processing System
STM	Serpentine Truss Manipulator
VGT	Variable Geometry Truss

II. PROTOTYPE PIPS HARDWARE

2.1 MODIFIED FOSTER-MILLER HARDWARE

Prototype PIPS hardware was built for NASA by Foster-Miller, Inc. The original hardware consisted of a motion base and serpentine truss manipulator (STM), with eight degrees-of-freedom (dof) [3]. Two "dummy boxes" and several static members were included in place of planned actuators. NASA has modified the original hardware significantly to give the prototype good serpentine capability. The "dummy boxes" and static members were replaced with active elements to give the system eighteen-dof.

The motion base has two-dof, a ball-screw track (prismatic joint) plus a pitch joint (crane joint). The kinematic diagram for the motion base is shown in Fig. 2.1. The primatic joint variable is d_1 and the pitch joint angle is θ_2. Frame {0} is the base coordinate frame for the system. Frame {3} is aligned 45° with respect to the ball-screw track. Fig. 2.1 also shows the attachment and first two segments of the STM. Figure 2.1 shows the zero position for θ_2; nominal follow-the-leader trajectories have $\theta_2 = -90°$.

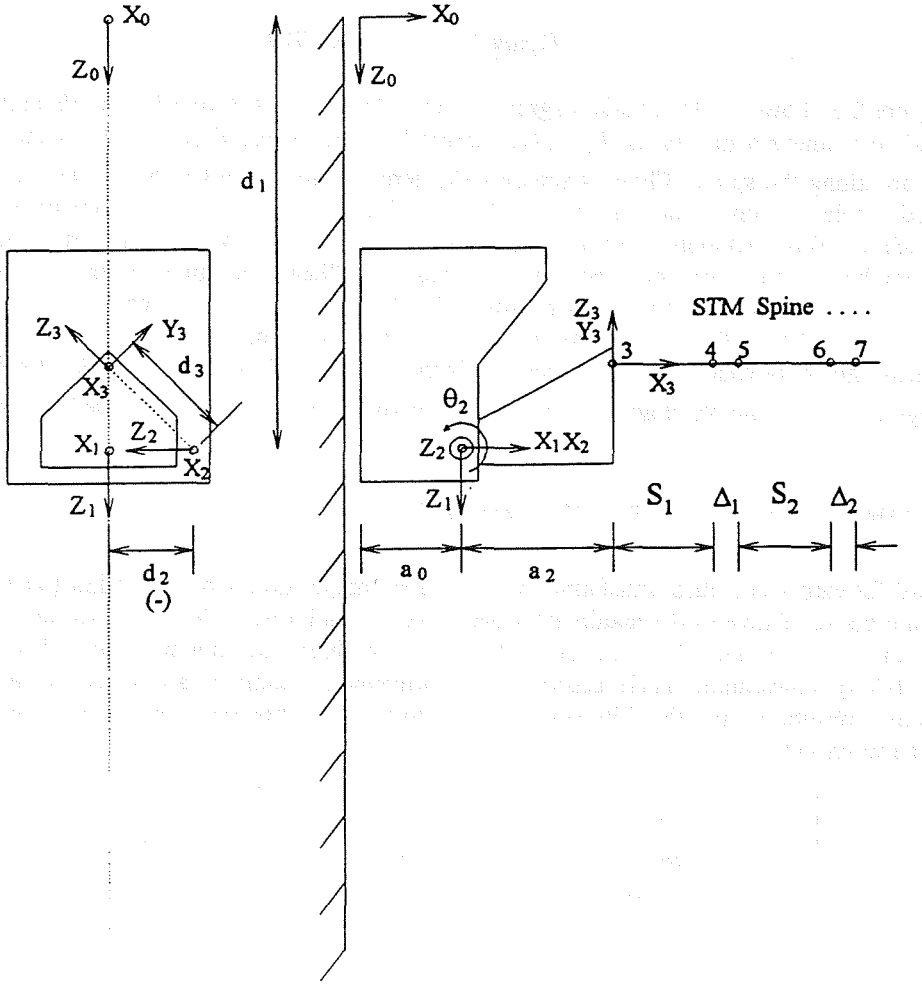


Figure 2.1 Motion Base for Prototype STM

The STM is a sixteen-dof tetrahedral variable geometry truss (VGT) as shown in Fig. 2.2. The rotation axes are orthogonal, rather than spaced by 120° as in the standard tetrahedron VGT from the literature [10]. The corners of the solid tetrahedra along the manipulator from base to tip, opposite the linear actuators, are called the spine of the STM. Figure 2.2 is mirror-image of the as-built hardware. Figure 2.3 depicts the prototype hardware geometry.

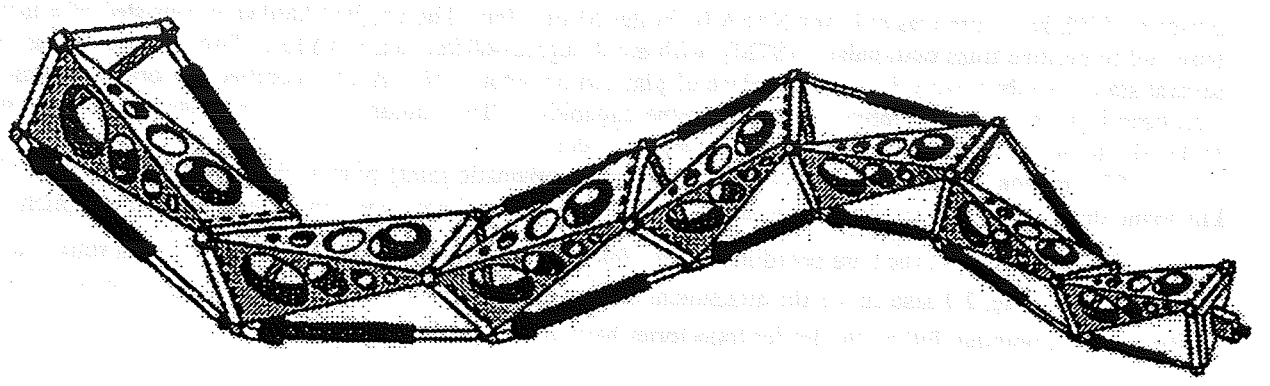


Figure 2.2 Prototype STM

Figure 2.3 shows the kinematic diagram for the STM. Views A and B are "flattened" about the spine. There are sixteen linear actuators $L_3, L_4, ..., L_{18}$, controlling the angles $\theta_3, \theta_4, ..., \theta_{18}$ about axes $Z_3, Z_4, ..., Z_{18}$. All X axes are along the spine. Figure 4 presents the zero position for all STM angles. The linear actuator to angular motion relationship is not as simple as shown in Fig. 2.3. The detailed geometry is documented in [3]. The Foster-Miller "Box" notation is given. This shows that the taper along the arm is discrete (three large boxes, three medium boxes, and two small boxes). An alternate "Segment" numbering for the STM modules is introduced in this report for generality, as shown. The Spine Points i are the origins of coordinate frames $\{i\}$. Spine Point 3 is attached to the motion base, and Spine Point 18 is associated with the last moving joint. The DH link lengths are given for each segment i, where S_i is the major length and Δ_i is the joint offset. Each joint pair bridging segment pairs is an offset universal joint. The module lengths Q_i are discussed in Section III.

2.2 FORWARD KINEMATICS TRANSFORMATIONS

The standard forward kinematics transformation calculates the position and orientation (pose) of the manipulator tip coordinate frame relative to the manipulator base given the joint variables and manipulator geometry. In this report, the motion base and STM are taken to be one eighteen-dof manipulator. The Denavit-Hartenberg parameters (Craig convention, [11]) describing the kinematic geometry are given in Appendix A. These parameters are substituted into the following equation to yield the homogeneous tranformation matrices relative neighboring frames [11].

$$ {}^{i-1}_{i}T = \begin{bmatrix} c\theta_i & -s\theta_i & 0 & a_{i-1} \\ s\theta_i c\alpha_{i-1} & c\theta_i c\alpha_{i-1} & -s\alpha_{i-1} & -d_i s\alpha_{i-1} \\ s\theta_i s\alpha_{i-1} & c\theta_i s\alpha_{i-1} & c\alpha_{i-1} & d_i c\alpha_{i-1} \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} & \left[{}^{i-1}_{i}R\right] & & \left\{{}^{i-1}P_i\right\} \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad (2.1) $$

As shown in Eq. 2.1, a homogeneous transformation matrix is composed of a rotation matrix representing the orientation and a position vector the translation between two frames. An artificial fourth row is added to maintain proper mathematical operations. Given all ${}^{i-1}_{i}T$, the forward kinematics transformation is:

$$ {}^{0}_{tip}T = {}^{0}_{1}T\,{}^{1}_{2}T\,{}^{2}_{3}T \ldots {}^{17}_{18}T\,{}^{18}_{tip}T , \tag{2.2} $$

where:

$$ {}^{18}_{tip}T = \begin{bmatrix} & [I] & & \{{}^{18}P_{tip}\} \\ 0 & 0 & 0 & 1 \end{bmatrix} . \tag{2.3} $$

Computation of the forward kinematics transformation is accomplished recursively, and yields a unique result. The inverse kinematics solution and graphical animation presented later requires intermediate rotation matrices and position vectors, respectively, obtained from:

$$ {}^{0}_{k}T = {}^{0}_{1}T\,{}^{1}_{2}T\,{}^{2}_{3}T \ldots {}^{k-2}_{k-1}T\,{}^{k-1}_{k}T , \qquad 1 \le k \le 18 . \tag{2.4} $$

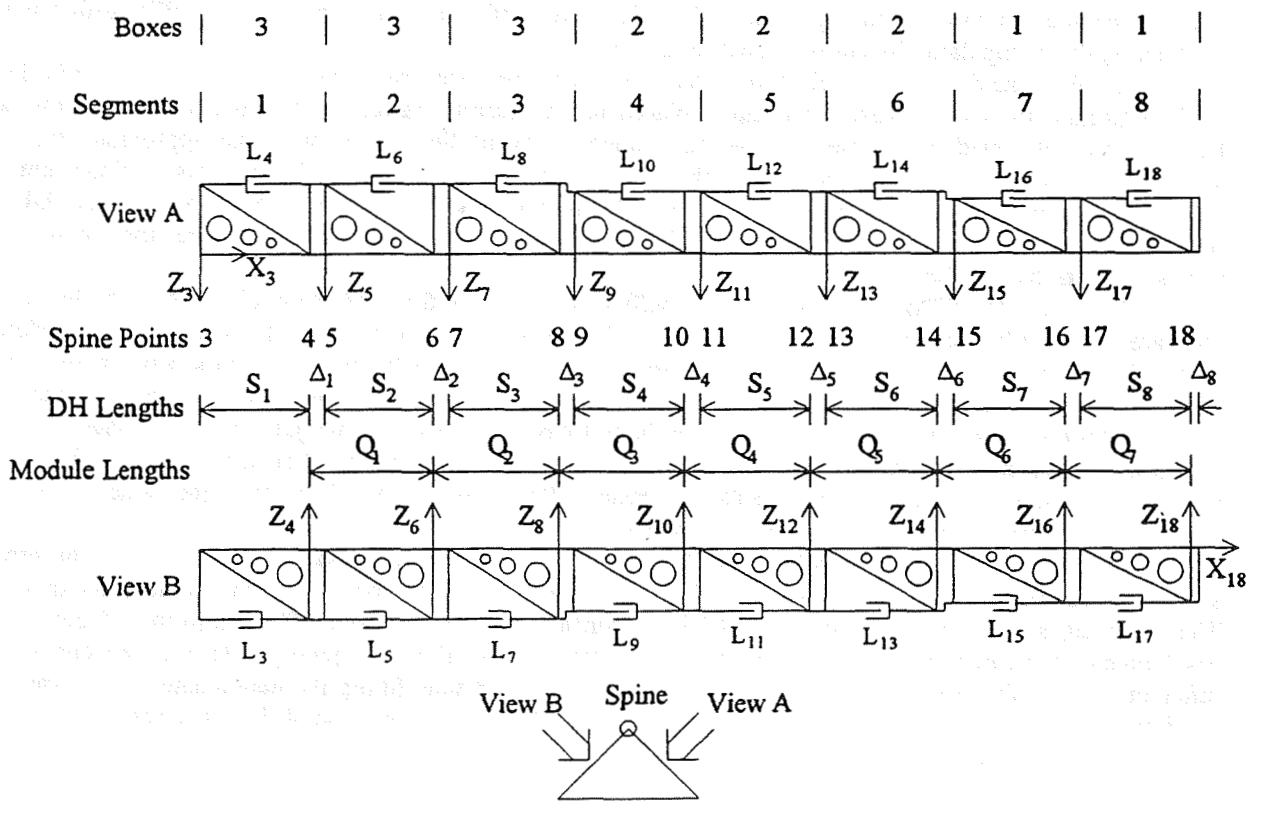


Figure 2.3 Kinematic Diagram for Prototype STM

III. FOLLOW-THE-LEADER ALGORITHM

3.1 GENERAL FOLLOW-THE-LEADER ALGORITHM

A general, modular follow-the-leader (FTL) algorithm for hyper-redundant serpentine manipulators was developed and implemented in computer simulation. It was used to simulate FTL motion for the prototype PIPS hardware. Given an obstacle-free trajectory for the manipulator tip, the FTL algorithm ensures obstacle-free motion for the entire manipulator by forcing ensuing links to follow the tip link.

The developed FTL algorithm is based on an extension of the method in [3]. In that work, the serpentine manipulator had fewer freedoms; the algorithm has been extended to include serpentine motion for each manipulator segment and a general treatment for the motion base. The current algorithm is adaptable to different manipulator and motion base kinematics with a minimum of code changes.

The FTL algorithm moves the manipulator tip along the given trajectory from the start to the end. The trajectory is described by a set of *XYZ* points, a piecewise linear discretization of the continuous trajectory. There is no limitation on the trajectory discretization compared to the size of segment lengths. The FTL algorithm controls the locations of spine points along the manipulator. A minimum of three dof are necessary in the motion base to place the base-most spine point at general 3D points. The motion base is used to feed the STM into the trajectory. Two dof are required to position one point relative to another along the given trajectory and separated by a fixed link distance. If the STM is composed of true universal joints (no offset between the orthogonal joint pairs) every spine point can be placed on the trajectory. For general STMs consisting of offset revolute joints, only every other spine point can be placed on the trajectory. The FTL algorithm controls only the STM spine; therefore, the path planning algorithm must provide trajectories with sufficient clearance to allow collision-free motion considering the manipulator dimensions about the spine.

To place the STM tip at a given trajectory point, FTL performs two steps: 1) The manipulator is shaped to the trajectory from tip to base. The tip is placed on the current trajectory point and in-board spine point locations are calculated by intersecting the manipulator segment link sphere with the appropriate trajectory straight-line segment. Each solution becomes the sphere center for the next link. The process continues until the base-most spine point is placed on the trajectory. For this purpose, the trajectory is appended with the straight line from the fixed base to the first trajectory point. 2) Inverse position kinematics then calculates the required joint values, from the base to the tip.

For zero-offset STMs the above steps are sufficient. However, for STMs with joint offsets, the link sphere radii are not fixed but functions of the intermediate joint angles. Therefore, steps 1) and 2) must be performed iteratively (the process starts with an average value for link radii) until the position errors between the desired spine points on the trajectory and the actual spine points achieved by inverse kinematics are sufficiently small.

This process is repeated for each point on the trajectory. At each step, the joint values are saved. Smooth serpentine motion may be obtained by providing a fine trajectory discretization. For operation, the off-line joint values are downloaded to the real-time STM controller. Retraction of the STM along the same obstacle-free trajectory is accomplished by reversing the joint values array.

Figure 3.1 gives a flow chart for the FTL process. The inner iteration loop drives the spine points error to a user-defined tolerance, with a maximum number of iterations. The outer loop drives the tip along the trajectory. The following sections present the detailed FTL algorithm for the prototype PIPS hardware. However, the development is modular so that only specific manipulator kinematics and spine points must be changed for different STMs. FTL derivations are presented for trajectory generation, fitting the manipulator to the trajectory, and inverse kinematics solutions. The FTL discussion concludes with presentation of simulated results.

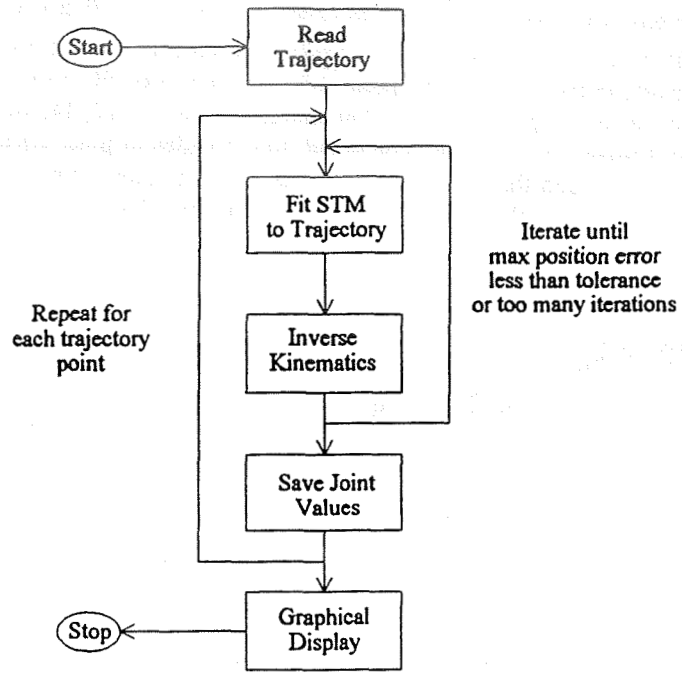


Figure 3.1 Follow-the-Leader Flowchart

3.2 TRAJECTORY GENERATION

This step will eventually be accomplished with a path planning algorithm. Currently, an obstacle-free trajectory for the manipulator tip is determined manually and discretized by computer. Four trajectories were developed to demonstrate the FTL algorithm. These were developed relative to a shuttle pallet and generic cylindrical payload to display varied motion.

The laboratory hardware is currently arranged so the prismatic joint is vertical, as shown in Figure 2.1. This should be changed so that the prismatic joint is horizontal, feeding the STM into a vertical Shuttle bay pallet. For this case, the motion base crane joint value would be $\theta_2 = -90^\circ$ so that the STM is aligned with the prismatic joint direction. The home position for the STM is the minimum d_1, with $\theta_3 = \theta_4 = ... = \theta_{18} = 0$. The trajectory must start at the manipulator tip in this home position. The middle and end trajectory points must be reachable subject to joint limits.

3.3 FIT STM TO TRAJECTORY

This section presents the first basic FTL component, fitting the hyper-redundant manipulator to the given trajectory. The method is general and will be applied to the specific hardware in Section II.

3.3.1 SPINE POINTS

Given an STM to control in FTL mode, the first step is to determine which points to control as the spine points. Since three dof are required to position a point independently in 3D space, the first serpentine joint, θ_3, is

combined with the motion base variables, d_1 and θ_2 in order to position the first spine point 4. Spine point i is the origin of coordinate frame $\{i\}$. Since two dof are required to position a point a given distance from another point, along a given trajectory, every second spine point following 4 can be fit to the trajectory. Therefore, there are a total of eight spine points for the prototype PIPS hardware: 4, 6, 8, 10, 12, 14, 16, and 18. In this paradigm, STM joint values from neighboring segments are combined into modules to place each ensuing spine point. The motion base is pictured in Fig 2.1 and the STM in Figs. 2.2 and 2.3. Figure 3.2 shows the spine of general STM module i controlling spine point j+2 with respect to spine point j using STM joint angles θ_{Ei} and θ_{Oi} (E and O stand for even and odd, respectively).

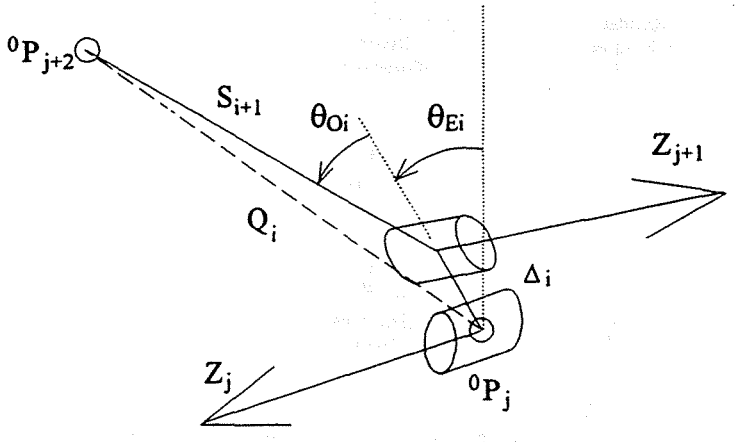


Figure 3.2 General Prototype STM Module

The intermediate spine point j+1 cannot be placed in general on the given trajectory. The fixed DH lengths are Δ_i and S_{i+1}, from segments i and i+1, respectively. The DH parameters for Module i are given in Table 3.1 (from Table A.1).

Table 3.1 DH Parameters for Module *i*

joint	α_{i-1}	a_{i-1}	d_i	θ_i
j	-90	S_i	0	θ_{Ei}
$j+1$	90	Δ_i	0	θ_{Oi}
$j+2$	-90	S_{i+1}	0	θ_{Ei+1}

The variable distance Q_i from spine points j to j+2 is the magnitude of the vector $^jP_{j+2}$; it is a function of the intermediate joint angle θ_{Oi}, as derived by forward kinematics.

$$\begin{Bmatrix} {}^jP_{j+2} \\ 1 \end{Bmatrix} = {}^j_{j+1}T \begin{Bmatrix} {}^{j+1}P_{j+2} \\ 1 \end{Bmatrix} = {}^j_{j+1}T(\theta_{Oi}) \begin{Bmatrix} S_{i+1} \\ 0 \\ 0 \\ 1 \end{Bmatrix} \tag{3.1}$$

$$ {}^{j}_{j+1}T = \begin{bmatrix} c\theta_{Oi} & -s\theta_{Oi} & 0 & \Delta_i \\ 0 & 0 & -1 & 0 \\ s\theta_{Oi} & c\theta_{Oi} & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3.2} $$

$$ {}^{j}P_{j+2} = \begin{Bmatrix} \Delta_i + S_{i+1}c\theta_{Oi} \\ 0 \\ S_{i+1}s\theta_{Oi} \end{Bmatrix} \tag{3.3} $$

$$ Q_i(\theta_{Oi}) = \left\| {}^{j}P_{j+2} \right\| = \sqrt{\Delta_i^2 + 2\Delta_i S_{i+1}c\theta_{Oi} + S_{i+1}^2} \tag{3.4} $$

Equation 3.4 is used to update the seven module lengths when iterating to reduce the maximum spine point position error (see Fig. 3.1). The Q_i are updated after the inverse kinematics solution.

Table 3.2 summarizes the spine point paradigm for the prototype hardware, giving module number, segments composing each module, spine points of the module end and start, fixed DH lengths, joint variables, and variable module length. The first row of Table 3.2 gives the motion base information; the seven STM modules follow.

Table 3.2 STM Modules

Module *i*	Segments	Spine Points	DH Lengths	Joint Variables	ModuleLength
-	Motion Base/1	4 / 0	-	d_1, θ_2, θ_3	-
1	1 / 2	6 / 4	Δ_1, S_2	θ_4, θ_5	Q_1
2	2 / 3	8 / 6	Δ_2, S_3	θ_6, θ_7	Q_2
3	3 / 4	10 / 8	Δ_3, S_4	θ_8, θ_9	Q_3
4	4 / 5	12 / 10	Δ_4, S_5	θ_{10}, θ_{11}	Q_4
5	5 / 6	14 / 12	Δ_5, S_6	θ_{12}, θ_{13}	Q_5
6	6 / 7	16 / 14	Δ_6, S_7	θ_{14}, θ_{15}	Q_6
7	7 / 8	18 / 16	Δ_7, S_8	θ_{16}, θ_{17}	Q_7

In this paradigm, the last STM joint angle θ_{18} is not required to place the last spine point, 18. It can be used in conjunction with camera pointing. For the prototype STM hardware, the following relates module index i and spine point index j: $j = 2(i+1)$.

3.3.2 INTERSECT ROUTINE

3.3.2.1 The Process

To fit a general STM to a given trajectory, the following process is used. The manipulator tip (the last spine point) is placed on the current trajectory point. The next in-board spine point is fit to the trajectory by intersecting a sphere centered at the last spine point and radius equal to the module length with the trajectory. The module lengths initialize at their average values (see Table A.2) and are updated with the current joint angles after each inverse kinematics solution. The intersection attempt starts at the trajectory line segment containing the current trajectory point. If no intersection is found, preceding trajectory segments are tried until the intersection is found. For ensuing modules, the newly calculated spine point becomes the sphere center and the next in-board module length the radius. This process repeats until the first spine point is placed on the trajectory.

This process is pictured for the prototype STM hardware in Fig. 3.3. In addition to the basic process, detail for module 6 is shown to demonstrate that the intermediate point 15 cannot be placed on the trajectory.

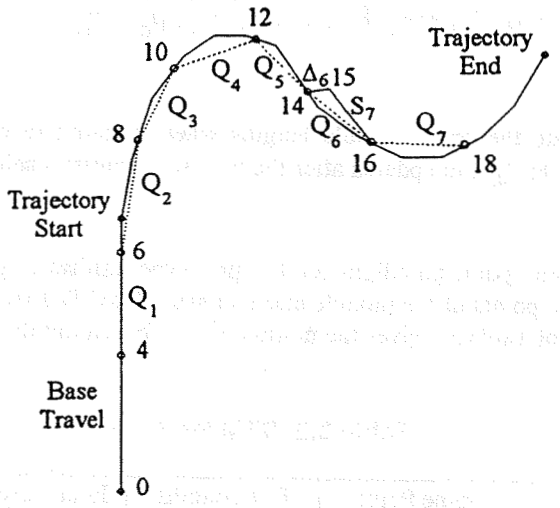


Figure 3.3 Fit STM to Given Trajectory

3.3.2.2 Intersection of Sphere with Line Segment

For the basic process of fitting the STM to a given trajectory described above, the intersection of a sphere with a 3D line segment must be solved repeatedly. This section presents the mathematics for this solution. The author thanks Sean for a fruitful discussion in this area.

The parametric form for a spatial line from point P_1 to point P_2 is:

$$\begin{aligned} x &= at + P_{1x} & a &= P_{2x} - P_{1x} \\ y &= bt + P_{1y} & b &= P_{2y} - P_{1y} \\ z &= ct + P_{1z} & c &= P_{2z} - P_{1z} \end{aligned} \qquad \begin{aligned} t &= 0 @ P_1 \\ t &= 1 @ P_2 \end{aligned} \tag{3.5}$$

The equation of a sphere with radius r and center x_c, y_c, z_c is:

$$(x - x_c)^2 + (y - y_c)^2 + (z - z_c)^2 = r^2 \tag{3.6}$$

The intersection of the sphere surface and line segment is found by substituting *x,y,z* from Eq. 3.5 into Eq. 3.6, yielding:

$$At^2 + Bt + C = 0 \tag{3.7}$$

$$\text{where: } \begin{aligned} A &= a^2 + b^2 + c^2 & d &= P_{1x} - x_c \\ B &= 2(ad + be + cf) & e &= P_{1y} - y_c \\ C &= d^2 + e^2 + f^2 - r^2 & f &= P_{1z} - z_c \end{aligned} \tag{3.8}$$

Solve t using the quadratic formula, $t_{1,2} = \dfrac{-B \pm \sqrt{B^2 - 4AC}}{2A}$; then evaluate *x,y,z* from Eqs. 3.5. There are two solutions, easily seen by imagining a pencil passed through a softball. The following intersection conditions exist for each of the two solutions.

- If $0 < t_i < 1$, the intersection lies on the line segment.
- If $t_i < 0$ or $t_i > 1$ the intersection lies on the line but off the line segment.
- If $t_1 = t_2$ the one intersection occurs where the line is tangent to the sphere; the above two rules apply.
- If $t_{1,2}$ is imaginary, no intersection exists.

3.4 INVERSE POSITION KINEMATICS SOLUTION

Given an STM fit to a given trajectory, this section presents the second basic FTL component, inverse position kinematics. The solution of this problem yields the joint values given the spine points. Only *XYZ* positions are satisfied, because the trajectory must give the nominal orientation at the STM tip. The intermediate orientations are used in the solution but not controlled independently. First, the motion base solution is presented, followed by the general solution which applies to each STM module *i*.

3.4.1 MOTION BASE SOLUTION

The motion base inverse position problem is: Given ${}^0P_4$, calculate d_1, θ_2, θ_3. The associated DH parameters are the first three lines in Table A.1. The kinematic diagram is shown in Fig. 2.1. Equation 3.9 gives the transformation relating the variables to the given information.

$${}^0P_4 = {}^0_3T\,{}^3P_4 = {}^0_1T(d_1)\,{}^1_2T(\theta_2)\,{}^2_3T(\theta_3)\begin{Bmatrix} S_1 \\ 0 \\ 0 \end{Bmatrix} \tag{3.9}$$

This equation expands to Eq. 3.10, where ${}^0P_4$ is given.

$${}^0P_4 = \begin{Bmatrix} X \\ Y \\ Z \end{Bmatrix} = \begin{Bmatrix} a_0 + c\theta_2(S_1 c\theta_3 + a_2) - Ks\theta_2(S_1 s\theta_3 + d_3) \\ d_2 - K(S_1 s\theta_3 - d_3) \\ d_1 - s\theta_2(S_1 c\theta_3 + a_2) - Kc\theta_2(S_1 s\theta_3 + d_3) \end{Bmatrix} \qquad K = \frac{\sqrt{2}}{2} \tag{3.10}$$

The solution of Eq. 3.10 follows the order: 1) θ_3; 2) θ_2; 3) d_1. From the Y component of Eq. 3.10:

$$\theta_3 = \sin^{-1}\left(\frac{d_2 + Kd_3 - Y}{KS_1}\right) \tag{3.11}$$

There is a unique solution: considering joint limits, the inverse sine ambiguity presents no problem. Given the θ_3 solution, the X component of Eq. 3.10 yields:

$$E\cos\theta_2 + F\sin\theta_2 + G = 0 \qquad \begin{aligned} E &= S_1 c\theta_3 + a_2 \\ F &= -K\left(S_1 s\theta_3 + d_3\right) \\ G &= a_0 - X \end{aligned} \tag{3.12}$$

Equation 3.12 is solved using the tangent half-angle substitution:

$$t = \tan\left(\frac{\theta}{2}\right) \quad \cos\theta = \frac{1-t^2}{1+t^2} \quad \sin\theta = \frac{2t}{1+t^2} \tag{3.13}$$

Substituting Eq. 3.13 into Eq. 3.12, simplifying, solving t with the quadratic formula, and using $t = \tan\left(\frac{\theta}{2}\right)$ yields two valid solutions for θ_2, given in Eq. 3.14. The solution chosen for FTL control is the value closest to the nominal STM home position, $\theta_2 = -90°$.

$$\theta_{2_{1,2}} = 2\tan^{-1}\left(\frac{-F \pm \sqrt{E^2 + F^2 - G^2}}{G - E}\right) \tag{3.14}$$

Given the θ_3 and θ_2 solutions, the Z component of Eq. 3.10 solves the prismatic joint variable. There is one solution for each θ_2, θ_3 pair.

$$d_1 = Z + s\theta_2\left(S_1 c\theta_3 + a_2\right) + Kc\theta_2\left(S_1 s\theta_3 + d_3\right) \tag{3.15}$$

3.4.2 MODULE *i* SOLUTION

When the solution for the first three joint variables are known, the remaining STM joint angles can be found, proceeding from the base to the manipulator tip. The prototype STM hardware consists of repeating modules, as discussed in Section 3.3.1. Though the STM is tapered so the DH lengths are not repeating, the kinematic structure of the solution is identical for each module. The general solution derived below is applied seven times, from module one through seven. Since θ_{18} is not required to position the STM tip, spine point 18, it is set to zero, and can be used in conjunction with camera pointing after the serpentine motion is complete.

The general module inverse kinematics problem is: Given two consecutive spine point locations, plus the previous joint values, calculate θ_{Ei}, θ_{Oi}. The vector difference between neighboring spine points can be expressed in two ways:

$${}^0P_{j+2} - {}^0P_j = {}^0_jR\,{}^jP_{j+2} \tag{3.16}$$

Equation 3.16 is rearranged to show the dependence on the unknown joint angles.

$$ {}^{0}P_{j+2} - {}^{0}P_{j} = {}^{0}_{j-1}R\,{}^{j-1}_{j}R(\theta_{Ei})\,{}^{j}P_{j+2}(\theta_{Oi}) \tag{3.17} $$

The left-hand-side vectors are given from the known spine points, referred to the {0} frame. The rotation matrix ${}^{0}_{j-1}R = {}^{0}_{1}R\,{}^{1}_{2}R \ldots {}^{j-2}_{j-1}R$ is a known function of previously-determined joint angles.

$$ {}^{0}_{j-1}R^{T}\left({}^{0}P_{j+2} - {}^{0}P_{j}\right) = \begin{Bmatrix} x \\ y \\ z \end{Bmatrix} = {}^{j-1}_{j}R(\theta_{Ei})\,{}^{j}P_{j+2}(\theta_{Oi}) \tag{3.18} $$

The vector ${}^{j}P_{j+2}(\theta_{Oi})$ was given in Eq. 3.3. The rotation matrix expressing the unknown θ_{Ei} is given below:

$$ {}^{j-1}_{j}R(\theta_{Ei}) = \begin{bmatrix} c\theta_{Ei} & -s\theta_{Ei} & 0 \\ 0 & 0 & 1 \\ -s\theta_{Ei} & -c\theta_{Ei} & 0 \end{bmatrix} \tag{3.19} $$

Substituting these values into Eq. 3.18 yields:

$$ \begin{Bmatrix} x \\ y \\ z \end{Bmatrix} = \begin{Bmatrix} c\theta_{Ei}\left(\Delta_i + S_{i+1}c\theta_{Oi}\right) \\ S_{i+1}s\theta_{Oi} \\ -s\theta_{Ei}\left(\Delta_i + S_{i+1}c\theta_{Oi}\right) \end{Bmatrix} \tag{3.20} $$

In Eq. 3.20, the left-hand-side is known, while the right-hand-side contains the unknowns θ_{Ei}, θ_{Oi}. Solution of these unknown joint angles is obtained by equating like components. θ_{Ei} is solved from a ratio of the z to the x component. The four-quadrant inverse tangent function, *atan2*, must be used in Eq. 3.21.

$$ \frac{z}{x} = \frac{-s\theta_{Ei}\left(\Delta_i + S_{i+1}c\theta_{Oi}\right)}{c\theta_{Ei}\left(\Delta_i + S_{i+1}c\theta_{Oi}\right)}; \qquad \theta_{Ei} = a\tan 2(-z, x) \tag{3.21} $$

θ_{Oi} is solved from a ratio of the y to the x component; again, *atan2* is used.

$$ \frac{y}{\frac{x}{c\theta_{Ei}} - \Delta_i} = \frac{S_{i+1}s\theta_{Oi}}{S_{i+1}c\theta_{Oi}}; \qquad \theta_{Oi} = a\tan 2\left(y, \frac{x}{c\theta_{Ei}} - \Delta_i\right) \tag{3.22} $$

This completes the general solution for θ_{Ei}, θ_{Oi}. Once the motion base solution is known, this general solution is applied to the seven modules, substituting the appropriate lengths from Tables 3.2 and A.2. This process results in the values for joint angle pairs θ_4, θ_5 through θ_{16}, θ_{17}, from the formulas for θ_{Ei}, θ_{Oi}. As previously discussed, $\theta_{18} = 0$. Following the inverse kinematics solution, forward kinematics is used to calculate the errors between the desired and actual spine point locations. If the maximum error is too large, the module lengths are updated using the new joint angles and Eq. 3.4 and another iteration is made.

3.5 RESULTS

This section discusses FTL algorithm simulation results for the prototype PIPS hardware. This effort was accomplished using MATLAB, C++, and IGRIP.

3.5.1 MATLAB SIMULATION

MATLAB simulation software was used to develop, implement, and test the FTL algorithm. This code executes the steps in Fig. 3.1 for the prototype PIPS hardware, including graphical animation of the motion base and STM spine. Many trajectories were tested in the Shuttle pallet environment with a nominal cylindrical payload. Four trajectories (developed manually) are given in Appendix C.

The MATLAB code is modular and adaptable to different PIPS hardware designs by changing the forward and inverse kinematics modules. The code was delivered by floppy disk to Gabor Tamasi. Also on the disk are data files for the four trajectories (TRAJi.M, i = 1,2,3,4) and for the joint values results from the FTL algorithm (FTLi.M, i = 1,2,3,4). Appendix B presents more detail on the MATLAB code. This code and data is also available by calling the author at Ohio University, (614) 593-1096. Appendix D gives graphics resulting from the MATLAB FTL simulation, for trajectory four.

3.5.2 C++ IMPLEMENTATION

The author supported implementation of the FTL algorithm in C++ for the operational hardware.

3.5.3 IGRIP SIMULATION

A high-fidelity, kinematically-correct computer graphics model of the prototype PIPS hardware was developed using IGRIP software. Carey Cooper of Intergraph Corporation performed the IGRIP modeling based on information compiled by the author from reference [3], mechanical drawings of the prototype hardware, and the hardware itself. This simulation is used to demonstrate the FTL algorithm in a virtual laboratory. The four trajectories mentioned above were transferred to IGRIP. Joint angle histories, created by the MATLAB code for each trajectory, were downloaded to the simulation. A fictitious payload was added to the Shuttle pallet environment to simulate whole arm obstacle avoidance using the four trajectories and FTL algorithm. A videotape was produced to demonstrate the IGRIP FTL simulation. Figure D.5 shows the IGRIP model, where the STM is at the end of the fourth trajectory.

IV. CONCLUSION

4.1 Design Lessons Learned

When simulating motion of a robotic hardware system, one always learns lessons on how to improve the design. These issues are not always clear to the developers of hardware. Therefore, because the PCR PIPS hardware is to be developed soon, the following list of design lessons from the prototype hardware is presented.

- STM joint offsets should be zero.
- The motion base translational travel should be equal to the STM length.
- The simplest FTL algorithm results from equal STM link lengths.
- The FTL simulation should be used to design the final system kinematics.

The following list is more specific to the prototype hardware.

- The motion base must have more range in three dimensions.
- The first STM joints allow the STM to be driven into the prismatic track when $\theta_2 = -90^\circ$.
- Box 1 encoders do not allow full joint motion.
- The sensor skin also restricts joint motion.

4.2 Concluding Remarks

This report presents a summary of the author's summer 1995 work at NASA Kennedy Space Center. The general area is path planning for the Payload Inspection and Processing System (PIPS). Specifically, a follow-the-leader algorithm (FTL) was developed, implemented, tested, and demonstrated using computer graphics. The FTL algorithm ensures whole-arm collision avoidance for the PIPS manipulator, given an obstacle-free trajectory for the tip link. FTL limits the overall manipulator workspace, but the resulting reliable obstacle avoidance in a sensitive environment justifies the limitation.

The FTL algorithm is derived in general and implemented in modular code. This report focuses on the prototype PIPS hardware, but extension of the method to the PCR hardware requires changes only to the manipulator kinematics. In addition to the FTL algorithm documentation, this report gives kinematic information for the prototype hardware, derived and compiled from various sources. Extensive kinematics modeling, both forward and inverse, are presented to support the FTL algorithm.

Preliminary investigations were undertaken into the general path planning problem. The PIPS path planning algorithm will be accomplished in the IGRIP environment. There are two preliminary recommendations for determining obstacle-free trajectories for the camera at the manipulator tip: 1) Human-assisted path planning using IGRIP; 2) Automated path planning using built-in IGRIP functions. If existing algorithms solve the problem, there is no need to re-invent the wheel. The author proposes to return to implement path planning results to the hardware during summer 1996.

V. REFERENCES

[1] Richardson, B., Sklar, M., and Fresa, M., "PCR Inspection and Processing Robot Study, Final Report", McDonnell Douglas Space Systems - Kennedy Space Division, November, 1993.

[2] Pasch, K., "Self-Contained Deployable Serpentine Truss for Prelaunch Access of the Space Shuttle Orbiter Payloads", NAS-1659-FM-9106-387, Final Report, Contract NAS 10-11659, NASA Kennedy Space Center, FL, August, 1990.

[3] Snyder, M., "Self-Contained Deployable Serpentine Truss (SCDST) for Prelaunch Access of Space Shuttle Orbiter Payloads", NAS-1794-FM-9323-651, Final Report, Contract NAS 10-11794, NASA Kennedy Space Center, FL, October, 1993.

[4] Spanos, P.D., and Berka, R.B., "Development of a Large Space Robot: A Multi-Segment Approach - Part I", AIAA Paper AIAA-93-1463-CP, Proceedings of the 34th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, LaJolla, CA, April, 1993.

[5] Lee, T.S., Ohms, T., and Hayati, S., "A Highly Redundant Robot System for Inspection", AIAA Paper AIAA-94-1194-CP, Proceedings of the Conference on Intelligent Robotics in Field, Factory, Service, and Space (CIRFFSS '94), Houston, TX, March, 1994.

[6] Krogh, B.H., "A Generalized Potential Field Approach to Obstacle Avoidance Control", Proceedings of the SME Conference on Robotics Research, Bethlehem, PA, August, 1984.

[7] Lozano-Perez, T., "A Simple Motion Planning Algorithm for General Robot Manipulators", *IEEE Journal of Robotics and Automation*, Vol. RA-3, No. 3, June, 1987, pp. 224-238.

[8] Gupta, K.K., "Fast Collision Avoidance for Manipulator Arms: A Sequential Search Strategy", Proceedings of the IEEE International Conference on Automation and Robotics, 1990, pp. 1724-1729.

[9] Nakamura, Y., *Advanced Robotics*, Redundancy and Optimization, Addison-Wesley Publishing Co., Inc., Reading, MA, 1991.

[10] Subramaniam, M., and Kramer, S.N., "The Inverse Kinematic Solution of the Tetrahedron Based Variable-Geometry Truss Manipulator", *ASME Journal of Mechanical Design*, Vol. 114, September, 1992, pp. 433-437.

[11] Craig, J.J., *Introduction to Robotics: Mechanics and Control*, Addison-Wesley Publishing Co., Inc., Reading, MA, 1988.

APPENDIX A. DENAVIT-HARTENBERG PARAMETERS FOR THE PROTOTYPE PIPS HARDWARE

The Denavit-Hartenberg (DH) parameters (Craig convention, [11]) for the eighteen-dof prototype PIPS hardware are given in Table A.1. The units are degrees for angles and inches for lengths. The first three rows are for the motion base (Fig. 2.1), while rows four through eighteen represent the serpentine truss manipulator (Figs 2.2 and 2.3). The variables are d_1, the prismatic joint variable, and $\theta_2, \theta_3, \ldots, \theta_{18}$, the STM spine joint angles.

Table A.1 DH Parameters

i	α_{i-1}	a_{i-1}	d_i	θ_i
1	0	a_0	d_1	0
2	-90	0	d_2	θ_2
3	-45	a_2	d_3	θ_3
4	-90	S_1	0	θ_4
5	90	Δ_1	0	θ_5
6	-90	S_2	0	θ_6
7	90	Δ_2	0	θ_7
8	-90	S_3	0	θ_8
9	90	Δ_3	0	θ_9
10	-90	S_4	0	θ_{10}
11	90	Δ_4	0	θ_{11}
12	-90	S_5	0	θ_{12}
13	90	Δ_5	0	θ_{13}
14	-90	S_6	0	θ_{14}
15	90	Δ_6	0	θ_{15}
16	-90	S_7	0	θ_{16}
17	90	Δ_7	0	θ_{17}
18	-90	S_8	0	θ_{18}

The values for the general parameters in Table A.1 are given below.

$a_0 = 11.250$
$a_2 = 8.908$

$d_1 = \text{var.}$
$d_2 = -5.005$
$d_3 = 7.078$

$S_1 = S_2 = S_3 = 20.035$
$S_4 = S_5 = S_6 = 16.100$
$S_7 = S_8 = 12.000$

$\Delta_1 = \Delta_2 = 1.250$
$\Delta_3 = \Delta_6 = 1.500$
$\Delta_4 = \Delta_5 = 1.188$
$\Delta_7 = 0.813$
$\Delta_8 = 0.763$

Table A.2 gives specific data for the seven Modules i (Segment i-$i+1$) for the prototype STM; see Table 3.2. Modules 1,2 and 4,5 are identical. For each module, the DH lengths are given (see Table A.1). Also, information on the variable module length Q_i is included:

$$Q_{iAVG} = \frac{Q_{iMAX} + Q_{iMIN}}{2} \qquad \text{and} \qquad \Delta Q_i = Q_{iMAX} - Q_{iMIN}$$

are given for each module. Figure A.1, a planar view of Fig. 3.2, shows the definitions for Q_{iMAX} and Q_{iMIN}. The last column in Table A.2 gives the values of θ_{Oi} corresponding to Q_{iAVG}. Equation 3.4 was used for the last three columns of Table A.2.

Table A.2 Prototype STM Module Data

Module i	Segment i-$i+1$	Δ_i	S_{i+1}	Q_{iAVG}	ΔQ_i	θ_{OiAVG}
1,2	1-2,2-3	1.250	20.035	21.206	0.158	±21.1°
3	3-4	1.500	16.100	17.508	0.185	±21.1°
4,5	4-5,5-6	1.188	16.100	17.214	0.149	±21.1°
6	6-7	1.500	12.000	13.410	0.180	±21.1°
7	7-8	0.813	12.000	12.762	0.102	±21.1°

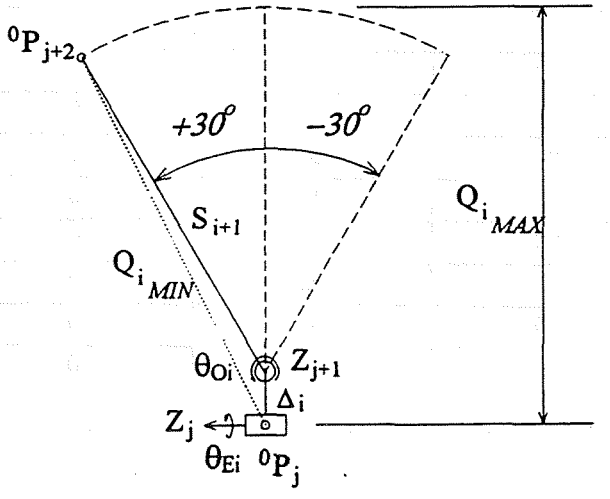


Figure A.1 STM Module i

The range of the prismatic joint, derived from mechanical drawings, is:

$$24.9" \le d_1 \le 114.5".$$

Nominal limits for all joint angles are ±30°. More detailed limits are published in [3]. Precise limits for the NASA-modified STM must be determined.

APPENDIX B. MATLAB CODE FOR THE FOLLOW-THE-LEADER ALGORITHM

MATLAB code for the follow-the-leader algorithm was delivered to Gabor Tamasi at the conclusion of the fellowship period. The format was ASCII m-files on a WINDOWS floppy disk. This code was also delivered and explained to Mike Sklar of McDonnell Douglas, who is developing a graphical user interface for PIPS. The FTL flowchart was given in Figure 3.1. MATLAB is an interpretive language. The main routine is:

FTL.M.

Additional m-files called from this main routine are:

PALPAY.M,
TRAJ1.M, TRAJ2.M, TRAJ3.M, TRAJ4.M,
INTRSECT.M,
INVKIN.M,
STMPLOT.M.

MATLAB functions, with their calling routines in parentheses, are:

SPHRLINE.M (INTRSECT.M),
DHFUN.M (INVKIN.M, STMPLOT.M),
TANHALF.M (INVKIN.M).

Variables used throughout the code:

traj	Trajectory points to follow.
pt	Point array from intersect routine.
alp, a, d, th	DH parameters.
d1, th2, th3, . . ., th18	Joint values from inverse kinematics.
jnt	Above joint values in one array.

Throughout the code, the following notation is used, with reference to [11].

tji : ${}^{i}T_{j}$ Homogeneous transformation matrix giving the position and orientation of frame $\{j\}$ with respect to frame $\{i\}$, expressed in $\{i\}$ coordinates.

rji : ${}^{i}R_{j}$ Orthonormal rotation matrix giving the orientation of frame $\{j\}$ with respect to frame $\{i\}$, expressed in $\{i\}$ coordinates.

pji : ${}^{i}P_{j}$ Position vector from the origin of frame $\{i\}$ to the origin of frame $\{j\}$, expressed in $\{i\}$ coordinates.

APPENDIX C. TRAJECTORY AND FTL DATA FILES

This appendix presents trajectory data files developed for the follow-the-leader simulation. These data are calculated and inserted into the traj variable by TRAJ1.M, TRAJ2.M, TRAJ3.M, and TRAJ4.M, respectively. All data are given with respect to frame {0} coordinates. The joint values from the follow-the-leader algorithm corresponding to these trajectory data files are FTL1.DAT, FTL2.DAT, FTL3.DAT, and FTL4.DAT. All TRAJi.DAT and FTLi.DAT files were delivered with the MATLAB code as mentioned in Appendix B.

Table C.1 TRAJ1.DAT

i	X	Y	Z
1	16.255	0.000	0.000
2	16.255	0.000	174.902
3	16.255	0.000	179.902
4	16.255	0.000	184.902
5	16.255	0.000	189.902
6	16.255	-0.656	196.139
7	16.255	-2.594	202.104
8	16.255	-5.729	207.536
9	16.255	-9.926	212.196
10	16.255	-15.000	215.883
11	16.255	-20.729	218.434
12	16.255	-26.864	219.738
13	16.255	-33.136	219.738
14	16.255	-39.271	218.434
15	16.255	-45.000	215.883

Table C.2 TRAJ2.DAT

i	X	Y	Z
1	16.255	0.000	0.000
2	16.255	0.000	174.902
3	15.255	0.000	179.902
4	13.255	0.000	184.902
5	10.255	0.000	189.902
6	7.255	-0.656	196.139
7	4.255	-2.594	202.104
8	1.255	-5.729	207.536
9	-1.745	-9.926	212.196
10	-4.745	-15.000	215.883
11	-7.745	-20.729	218.434
12	-10.745	-26.864	219.738
13	-13.745	-33.136	219.738
14	-15.745	-39.271	218.434
15	-16.745	-45.000	215.883

Table C.3 TRAJ3.DAT

i	X	Y	Z
1	16.255	0.000	0.000
2	16.255	0.000	174.902
3	17.255	0.000	179.902
4	19.255	0.000	184.902
5	22.255	0.000	189.902
6	25.255	-0.656	196.139
7	28.255	-2.594	202.104
8	31.255	-5.729	207.536
9	34.255	-9.926	212.196
10	37.255	-15.000	215.883
11	40.255	-20.729	218.434
12	43.255	-26.864	219.738
13	45.255	-33.136	219.738
14	46.255	-39.271	218.434
15	46.255	-45.000	215.883

Table C.4 TRAJ4.DAT

i	X	Y	Z
1	16.255	0.000	0.000
2	16.255	0.000	174.902
3	17.209	0.000	182.149
4	20.006	0.000	188.902
5	24.456	0.000	194.701
6	30.255	0.000	199.151
7	37.008	0.000	201.948
8	44.255	0.000	202.902
9	51.761	-0.988	202.902
10	58.755	-3.885	202.902
11	64.761	-8.494	202.902
12	69.370	-14.500	202.902
13	72.267	-21.494	202.902
14	73.255	-29.000	202.902

APPENDIX D. TRAJECTORY 4 RESULTS

D.1 SHUTTLE PALLET, CYLINDRICAL PAYLOAD, AND TRAJECTORY 4

Figure D.1 shows the Shuttle environment modeled in MATLAB. The 3D trajectory four is shown, consisting of two quarter circles in orthogonal planes. A planar cross section of the Space Shuttle pallet and a cylindrical payload are also shown.

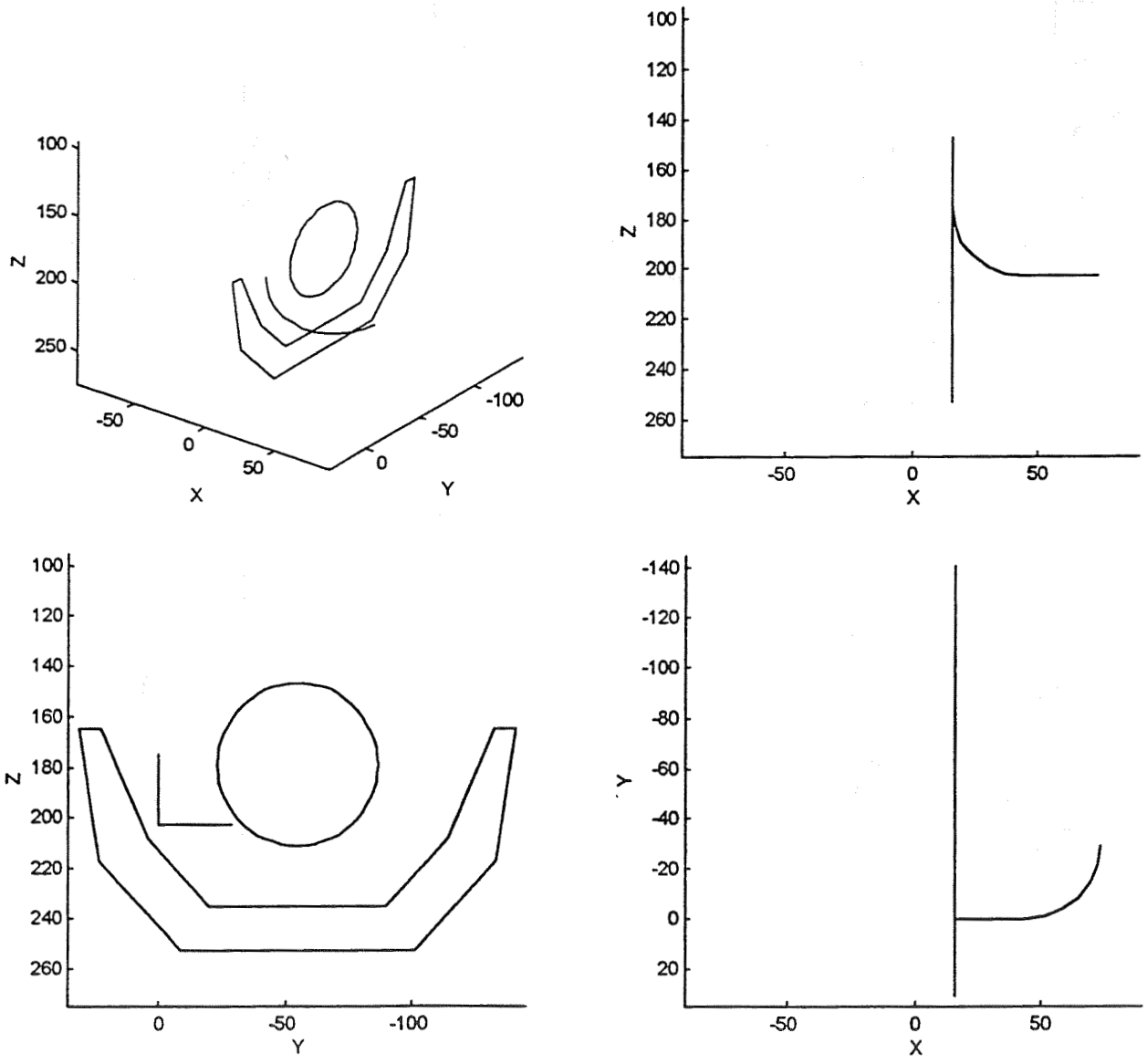


Figure D.1 Pallet, Payload, Trajectory 4

D.2 FOLLOW-THE-LEADER SIMULATION

Figures D.2 show the STM in FTL simulation. The figures progress from home position (Fig. D.2a) through two intermediate positions, and the end of trajectory four (Fig. D.2d). The appearance of interference of the STM and pallet is an illusion; trajectory four enables collision-free motion.

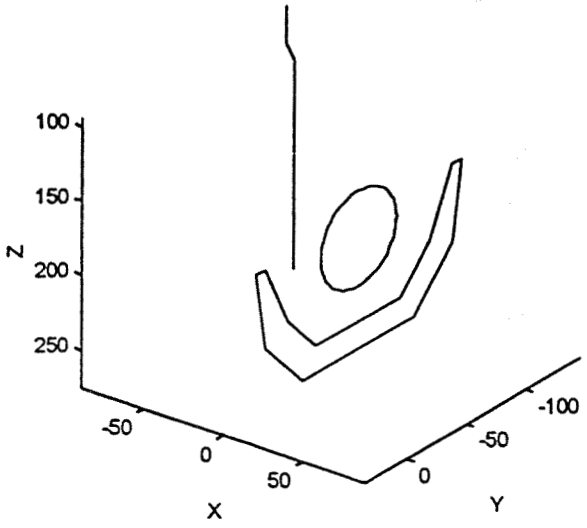


Figure D.2a

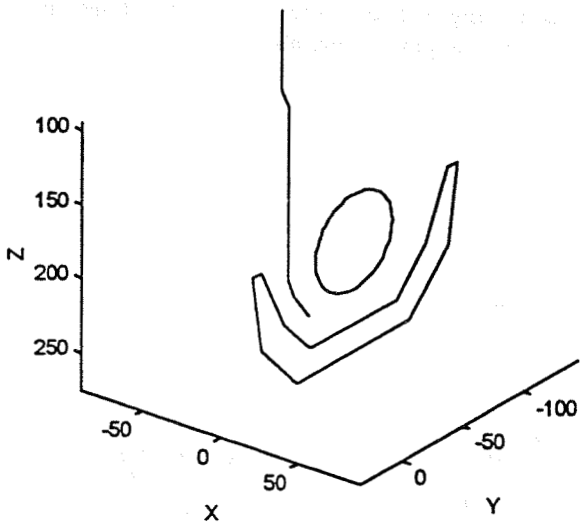


Figure D.2b

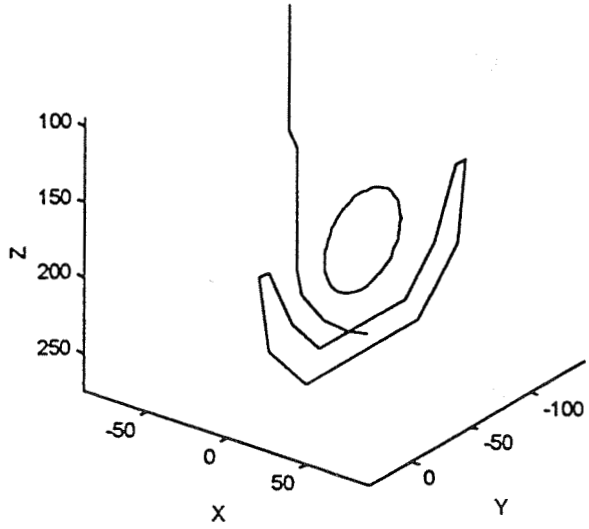


Figure D.2c

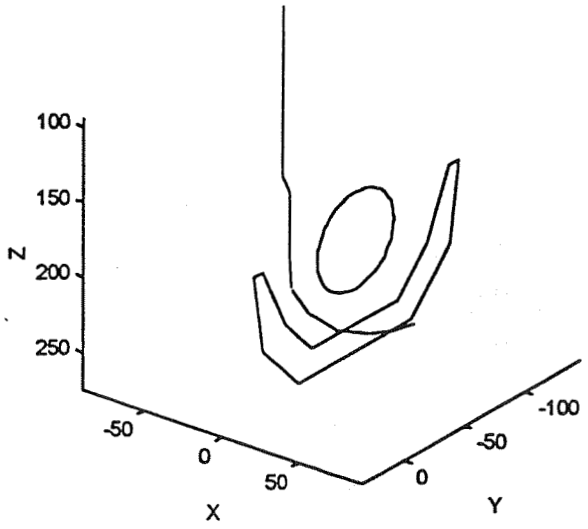


Figure D.2d

Figure D.2 FTL Simulation for Trajectory 4

D.3 JOINT VALUE RESULTS

Figures D.3 show the STM joint values from the FTL algorithm for trajectory four. The motion base values are shown, followed by the seven Module *i* values (Segments *i-i+1*).

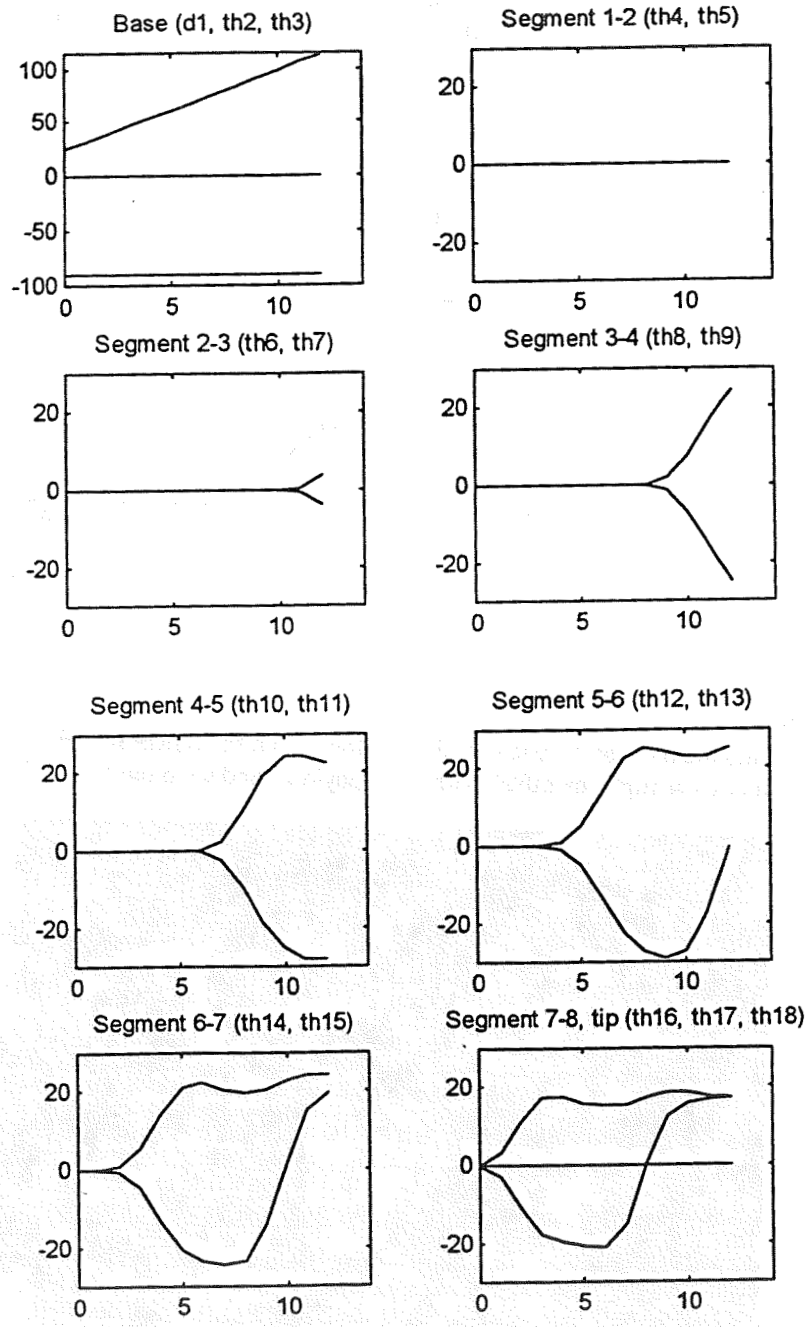


Figure D.3 FTL Joint Values

D.4 ITERATION RESULTS

Figures D.4 show the FTL algorithm convergence history for trajectory four. Two iterations are required for the first trajectory point, and one iteration thereafter to achieve an error tolerance of one-tenth of an inch (the horizontal line in Fig. D.4a). Since there is only one iteration after the initial step, the initial and final manipulator errors are identical. The error tends to increase as the manipulator traverses the trajectory. The error condition checks all spine points for convergence; the maximum for trajectory four was the tip error.

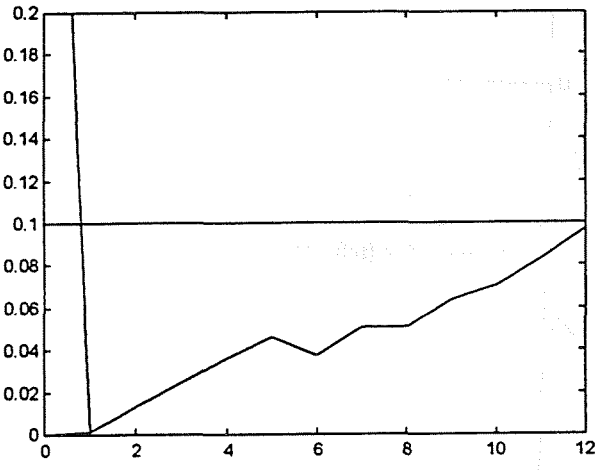


Figure D.4a FTL Tip Error

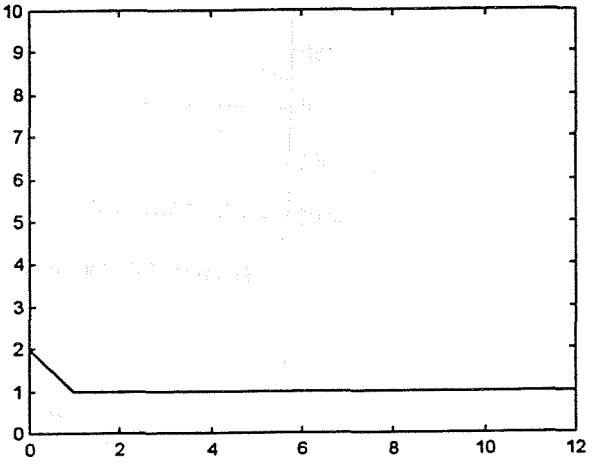


Figure D.4b FTL Iterations

D.5 IGRIP GRAPHICS

Figure D.5 shows the STM at the end of trajectory four, from the IGRIP model. This figure is the same as Fig. D.2d, with a different view angle, modified cylindrical payload, and solid model graphics.

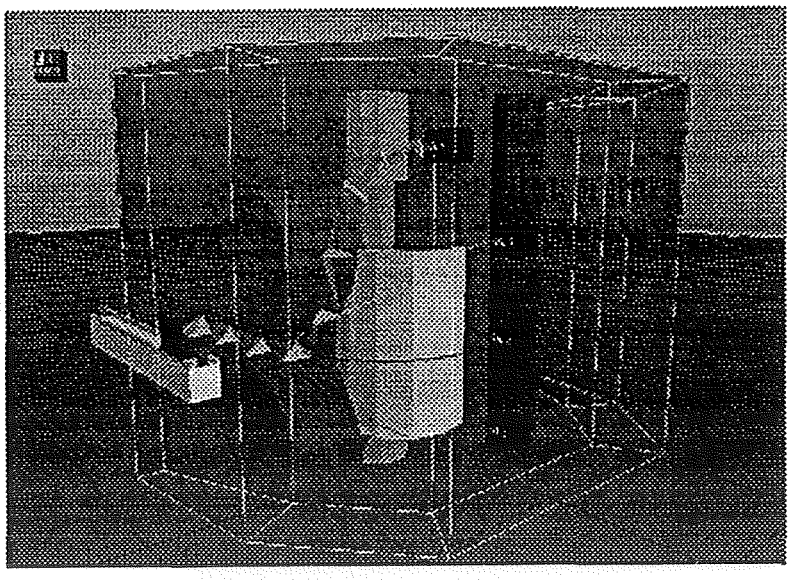


Figure D.5 IGRIP Model at Trajectory 4 End

ATTACHMENT - REPORT DOCUMENTATION PAGE - ITEM #6 (Authors)

1995 Research Reports

NASA/ASEE Summer Faculty Fellowship Program

REPORT AUTHORS:

Dr. Alfred S. Andrawis - South Dakota State University

Dr. William V. Brewer - Jackson State University (Mississippi)

Dr. John R. Cannon - University of Central Florida

Dr. Guillermo Colon - University of Puerto Rico-Mayaguez

Dr. Kemal Efe - University of Southwestern Louisiana

Dr. Gerald W. Evans - University of Louisville (Kentucky)

Dr. Gregory S. Forbes - The Pennsylvania State University

Dr. Raghvendra R. Gejji - Western Michigan University

Dr. Lucille A. Giannuzzi - University of Central Florida

Dr. Barbara H. Glasscock - California State Polytechnic University

Dr. Rhyn H. Kim - University of North Carolina-Charlotte

Dr. Gwendolyn D. Mitchell - Auburn University (Alabama)

Dr. Horacio A. Mottola - Oklahoma State University

Dr. Robert E. Peale - University of Central Florida

Dr. Rafael A. Perez - University of South Florida

Dr. Allen L. Rakow - Colorado State University

Dr. Dan E. Tamir - Florida Institute of Technology

Dr. Madjid Tavana - La Salle University (Pennsylvania)

Dr. Parveen F. Wahid - University of Central Florida

Mr. Pao-lien Wang - University of North Carolina-Charlotte

Dr. Jonathan E. Whitlow - Florida Institute of Technology

Dr. Robert L. Williams - Ohio University

EDITORS:

Dr. E. Ramon Hosler - University of Central Florida

Mr. Gregg Buckingham - John F. Kennedy Space Center